ASC 2003

The Impact of Technology on the Survey Process

Proceedings of the Fourth ASC International Conference

University of Warwick, UK, September 17-19, 2003

Edited by

Randy Banks
University of Essex

James Currall
University of Glasgow

John Francis
Office for National Statistics

Laurance Gerrard
TRBI

Raz Khan
ATP

Tim Macer
meaning ltd

Malcolm Rigg
BMRB

Ed Ross
Open Survey

Steve Taylor
RSM

Andrew Westlake
Survey & Statistical Computing

Association for Survey Computing

ASSOCIATION FOR SURVEY COMPUTING

PO Box 60, Chesham, Bucks, UK, HP5 3QH
tel/fax: +44 (0)1494 793 033
email: admin@asc.org.uk
http://www.asc.org.uk

ISBN: 0 9521682 9 4

British Library Cataloguing in Publication Data. A catalogue record for this book is available from the British Library.

Preface

Welcome to *ASC 2003, The Impact of Technology on the Survey Process*, the Association for Survey Computing's fourth major international conference on survey and statistical computing. As with others in the series, the aim of this one was to provide as broad a view as possible of the salient issues and challenges confronting the field.

Papers by three of the four speakers invited to address plenary sessions of the conference comprise the first section of this volume. Norman Glass, the Chief Executive of the National Centre for Social Research (NatCen) and our keynote speaker, tackles the main conference theme head-on, discussing the range of ways in which technological changes have impacted on NatCen's working practices over the past decade. David Pullinger, the Deputy Director Communication in the Office for National Statistics, examines the democratisation of official statistics in the UK and, in this context, the pull-push effect of web-technology, in particular, on the empowerment of civil society. Peter Mouncey, who, *inter alia*, is Director of Research at the Institute of Direct Marketing, considers both the threats posed and opportunities offered by new technologies to the reputation and independence of the survey research industry. Our end-note speaker, Ian Diamond, Director of the Economic and Social Research Council, was kind enough to step in rather too late in the day to produce a paper on his topic of *The Impact of Technology on the Survey Process. Future Prospects and the Role of the ESRC* in time for publication.[1]

The remainder and majority of this volume consist of contributed papers addressing one of the six major themes that emerged when the Scientific Programme Committee (SPC) grouped together the proposals that were accepted from amongst the many excellent offers we originally received. In the order in which they appear in this volume, they are:

- Organisational Behaviour
- Survey Development
- Survey Software
- Survey Modalities
- The Effectiveness of the Internet
- Data Dissemination

In some respects, these themes serve as headings for convenience only, as many papers could easily have fit under more than one heading. We expect the reader to find much pleasure in noting the many points of contact and convergence between the papers regardless of the immediate company they are seen to be keeping.

The contributed papers include both case studies and more theoretical reflections; some represent the outcome of purely methodological investigations, while others have a substantive focus; and those of us, shall we say, whose length of service now casts a significant shadow, will also find material on which quietly nostalgic reminiscences can be agreeably based. Regardless of content or technique,

[1] Ian replaces Denise Lievesley, Director of Statistics at Unesco, whose timetable for the period of the conference was unavoidably rescheduled.

however, we are sure that the reader will find the papers in this volume instructive, thought-provoking and, in many cases, directly and practically applicable to his or her own circumstances.

On behalf of the ASC we would like to thank all the authors for preparing what we feel is an excellent volume of papers and for putting up with our many urgent reminders about the deadlines we did not want them to overrun. As co-chairs of the SPC, we would also like to thank all of the members, in particular those who convened sessions and helped edit this volume and, in James Currall's case, went above and beyond the call of editorial duty. Thanks are also owing to Thomas Brennan, Peter Lynn, Jean Martin and Jennifer Waterton, ex-members of the SPC who were forced to resign for one reason or another. Finally, we would like to offer a very special Thank You to Rachel Sturge of *meaning ltd*, without whose extremely welcome efforts this publication would not have met its print-by date.

Randy Banks and Raz Khan, August 2003.

Members of the ASC 2003 Scientific Programme Committee

Randy Banks, *University of Essex* (Co-Chair)
Garj Basi, *BMRB*
James Currall, *University of Glasgow*
Steve Elder, *National Centre for Social Research*
John Francis, *Office for National Statistics*
Laurance Gerrard, *TRBI*
Raz Khan, *ATP* (Co-Chair)
Tim Macer, *meaning ltd*
Malcolm Rigg, *BMRB*
Ed Ross, *Open Survey*
Wendy Sykes, *Social Research Association*
Steve Taylor, *RSM*
Andrew Westlake, *Survey & Statistical Computing*

About the ASC

The Association for Survey Computing (ASC) is a not-for-profit organisation, based in the UK but drawing significant international membership and support. It provides a valuable forum for all those concerned with survey research and statistical computing and has a wide-ranging membership at both individual and corporate level. Originally formed in 1971 as the Study Group on Computers in Survey Analysis, it became the Association for Survey Computing in 1994. The ASC has links with the British Computer Society, the International Association for Statistical Computing, the Market Research Society, The Royal Statistical Society and the Social Research Association.

The ASC's many activities and links to other organisations are publicised through its website (http://www.asc.org.uk) and include the many conferences and workshops it runs, the register of available software it maintains and its support of other initiatives in the survey and statistical field.

For many years the ASC has organised one-day conferences in London in January and June together with occasional one-day specialised workshops. In 1992, to celebrate its 'coming of age', the ASC organised its first International 3-Day Conference, held in Bristol, England. Since then it has held similar international events at Imperial College in London, England (1996), at Edinburgh University, in Scotland (1999) and now, in 2003, holds its fourth International Conference here in Warwick.

The ASC was formed in order to improve knowledge of good practice in survey computing and to disseminate information on techniques and survey software. From the outset, a particular emphasis has been to help avoid wasteful duplication of computer programs where good packages already exist and to encourage the development of common standards. In its early days the focus of involvement was on collection and analysis of survey data. Now, after more than 30 years, the Association's role has extended much more widely from the involvement of computers in survey design and formulation at one end of the process to presentation and dissemination of survey data and results at the other. Over this period it has built up a unique role as an interchange point between a variety of survey and statistical specialisms and between client and contractor organisations.

The ASC maintains a Register of Software for Statistical and Survey Analysis, listing details of all relevant packages known to the Association. This is available on the ASC website and a printed version is also made available periodically to ASC members. It also publishes, for its membership, a periodic newsletter.

An important new ASC innovation, in co-operation with the Market Research Society is the ground-breaking 'MRS/ASC Award for Technology Effectiveness', intended to foster innovation in the involvement of computers in the survey process, whether on a large or small scale, by both companies and individual professionals. This award will be presented for the first time in the autumn of 2003 and the short-list of finalists will be announced at the ASC2003 Conference Dinner.

While principally dedicated to serving its individual and corporate membership, non-members are welcome at all events. Membership is inexpensive and brings a number of benefits. More details are available from the ASC's Administrator, PO Box 60, Chesham, Bucks, UK, HP5 3QH ('Phone/Fax - 01494-793033 [International - +44 1494-793033]; email admin@asc.org.uk).

The activities of the ASC are organised by a volunteer committee - elected annually - supported by an Administrator. The 2002/2003 ASC Committee was:

President:	Beverley Charles Rowe	Garj Basi *(BMRB)*
Chair:	Steve Elder *(NCSR)*	John Francis *(ONS)*
Vice-chair	Randy Banks (*University of Essex*)	Raz Khan *(ATP)*
Secretary	Andrew Westlake *(SASC)*	Tim Macer (*meaning ltd*)
Treasurer	Laurance Gerrard *(TRBI)*	Steve Taylor *(RSM)*
Administrator	Diana Elder	
Affiliates	*Suzanne Evans (RSS), Wendy Sykes (SRA), Sue Trenhaile (MRS)*	

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Table of Contents

Invited Speakers

Organisational Behaviour

Survey Development

Survey Software

Survey Modalities

The Effectiveness of the Internet

Data Dissemination

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

How Much Can Technology Help Research?

Norman Glass

Abstract

NatCen is the UK's largest independent social research organisation. The use of Computer-Assisted Interviewing in the early 90s made a dramatic change to the way research was carried out and to the productivity of our survey organisation. There has been no comparable technological breakthrough since and, for those wanting to carry out surveys with statistically representative samples of the population on complicated issues, the traditional survey issues of sample and questionnaire design, interviewer quality, response rates, accurate data and meaningful analysis still dominate and resist easy technological fixes.

Keywords

Computer assisted personal interviewing (CAPI); computer assisted telephone interviewing (CATI); computer assisted self interviewing (CASI); random Digit Dialling (RDD); probability sampling; quota sampling; response rate.

2. NatCen

The National Centre for Social Research (NatCen) is now the UK's largest independent social research organisation. It has 16 specialist project computing staff working in BLAISE with our 75 quantitative researchers, and a further 10 systems staff supporting our internal operations and fieldforce of 1000 interviewers and 200 nurse/interviewers. NatCen generated a research income of £23.5m in 2002/3, the vast bulk of it from projects involving computer-assisted interviewing. At any given moment around 30 projects are likely to be in the field with sample sizes ranging up to 25000 households and above (and many more individuals in total since we often interview all household members). NatCen interviews about 1% of all households in the country every year.

There are a number of distinguishing features of NatCen's work, of which not the least important is that we only use probability sampling to select our samples. NatCen believes strongly that the biases in quota sampling (its inherent tendency to under-represent difficult-to-contact households and individuals) can lead to seriously misleading results, for public policy at least, and, of course, render statistical analysis of variances problematic. Whilst response rates for quota sampling are rarely published (and indeed may not even have any strict meaning) it is clear that even, on their own terms, response rates to quota samples are often well below 20%. This compares with the response rates for most NatCen surveys of over 70%. This level of response rate is all the more striking when one considers the length of interview which is normal in most of NatCen's studies, all of which are carried out in a non-commercial context, the bulk being commissioned by UK government departments.

The core of NatCen's activity are four continuous annual surveys with samples ranging from 15000 to 25000 households. In the case of the Health Survey for England (carried out in collaboration with epidemiologists at UCL for the Department of Health) the interviews, which require interviewers to obtain some physical information, are supplemented by subsequent nurse visits to take blood samples and blood pressure measurements. The National Travel Survey, carried out for DfT, requires participants (every member of the household) to keep travel diaries for seven days, including the provision of detailed geographical information on one of the days.

NatCen's mission is to carry out high-quality and innovative work. Two recent examples are the Arrestees Survey for the Home Office, where we ask our interviewers to go into police stations and interview a random sample of arrestees shortly after they have been charged, in order to learn more about the role that drink and drugs play in their criminal behaviour; and the first large-scale randomised controlled trial of a social policy in the UK, the Job Retention and Rehabilitation Pilot, for DWP, in which employees who have been off work sick for some time volunteer to be randomly allocated between three experimental programmes and a control group in order to test the impact of the "treatment" programmes on their subsequent employment experience.

Chart 1 Average Length of Interview

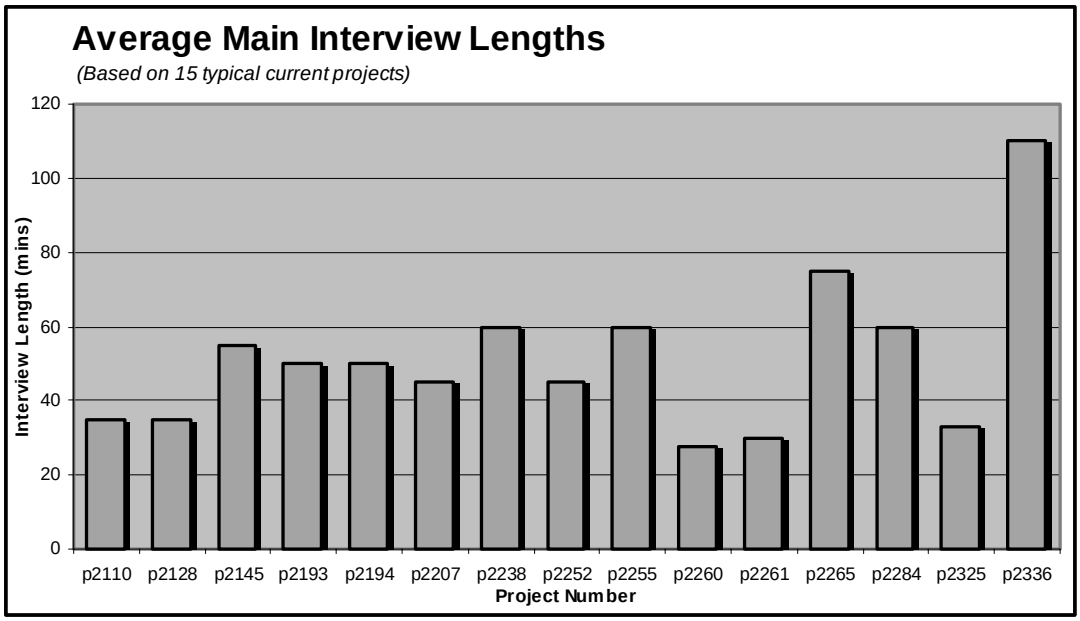


Chart 1 shows a typical sample of current NatCen surveys and the distribution of interview lengths in minutes. From this you will see that our standard interview length tends to be between 45 minutes and an hour but we have some interviews which can last as long as two hours (and more !). In general participants are not paid for their participation and this points up the importance of software which can cope with the complex branching in most of the surveys and yet be sufficiently interviewer friendly to allow interviewers to get through interviews quickly and competently in such a way as not to irritate or alienate the interviewee.

Much of NatCen's work, as I have said, consists of very large scale and very time-consuming nation-wide face-to-face surveys with pre-selected samples. To make it clear that, nevertheless, it is possible in such circumstances to achieve high response rates Table 1 shows a selection of some of our current large surveys with sample sizes and interview lengths compared with their response rates.

Table 1: NatCen Surveys: Samples, Response Rates and Interview lengths

	Sample Size	Response Rate	Interview Length (minutes)
Millennium Cohort	18500	82%	80
Health Survey	18000	75%	65
Survey of English Housing	20000	72%	30
Family Resources Survey	25000	66%	75
National Travel Survey	15000	69% (62% with all diaries complete)	90
Crime and Justice Survey	10000	75%	55

Achieving response rates of this kind for long and complex (and often very intimate) surveys of this kind with large samples, clearly requires careful planning, monitoring and organisation. Chart 2 (next page) sets out the typical stages of one of our surveys showing the lead times before we go into the field. We recognise, of course, that we are lucky to have clients who share our concern with the quality of the data we collect and are willing to pay the cost, and take the time, to get it.

It also requires a level of commitment to quality data and persistence among interviewers which literally has to be seen to believed. In some interviews respondents are expected to refer to recent payslips and bank statements and to search out documents relating to pensions, savings and other outgoings.

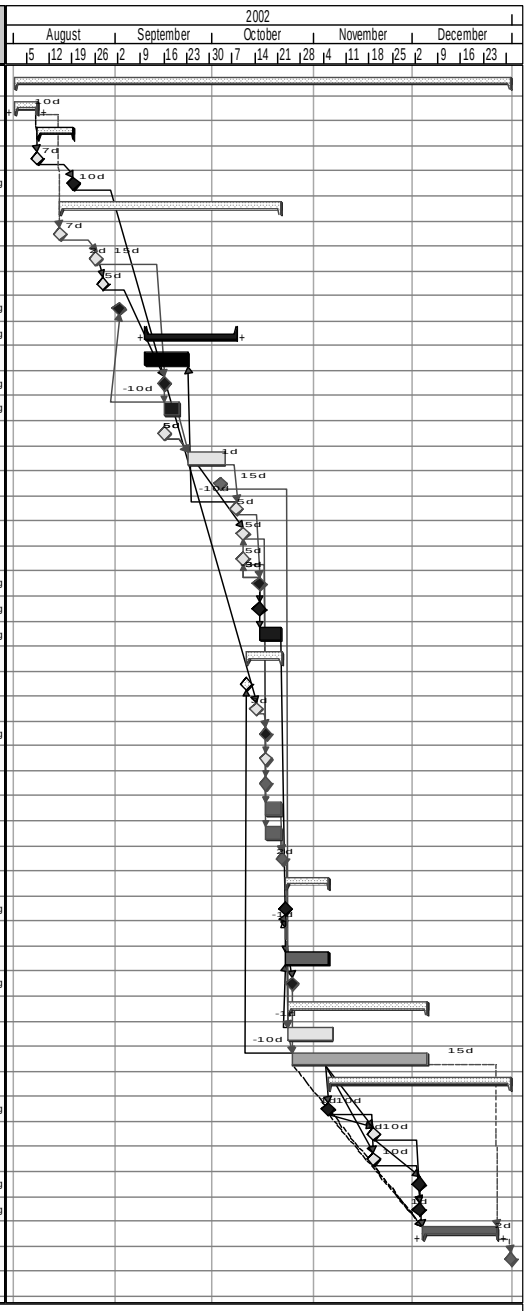
3. How can technology help

Following an abortive attempt at Computer Assisted Telephone Interviewing in the 80s NatCen successfully participated in the development and launch of large-scale Computer Assisted Personal Interviewing in the early 1990s with the Family Resources Survey, a survey of income, savings and wealth and employment for the then Department of Social Security in 1992. Coincidentally I was the Chief Economist in the Department at that time.

DSS needed a larger continuous annual survey than the existing 7000 households provided by the existing Family Expenditure Survey carried out by the Office of National Statistics to allow us to improve the accuracy of the poverty statistics, develop our tax-benefit policy modelling and to tell us more about important policy-relevant groups such as lone parents and disabled people. So we commissioned two consultants to advise us on the parameters of such a survey, with a specific remit to examine whether CAPI was feasible and desirable. NatCen (SCPR as it then was) and ONS were then commissioned to carry out the survey using CAPI to interview 25000 households on a continuous basis each year.

Chart 2: Timetable for a Typical Project

Line	Name	Duration	Start	End	Resource
1	**Whole Project**	**108d**	**01/08/02**	**01/12/02**	
2	**Project Start Up**	**5d**	**01/08/02**	**08/08/02**	
3	**Selection of Sample Points**	**7d**	**08/08/02**	**19/08/02**	
4	Finalise sample design		08/08/02	08/08/02	Researchers
5	Sample points selected & sent to AMs		19/08/02	19/08/02	Projects Computing
6	**Questionnaire Development and Pilot**	**48d**	**15/08/02**	**21/10/02**	
7	First list of questions agreed		15/08/02	15/08/02	Researchers
8	First CAPI specs		26/08/02	26/08/02	Researchers
9	Cut-off for CAPI pilot specs		28/08/02	28/08/02	Researchers
10	Set up install system for CAPI menu system		02/09/02	02/09/02	Projects Computing
11	Ongoing program development and testing	20d	10/09/02	08/10/02	Projects Computing
12	Period for client feedback on questionnaire	10d	10/09/02	23/09/02	Client
13	Pilot program ready for de-bugging		16/09/02	16/09/02	Projects Computing
14	De-bugging /testing/laptop performance testing of pilot program	5d	16/09/02	20/09/02	Projects Computing
15	Draft ARF circulated		16/09/02	16/09/02	Researchers
16	Pilot briefing / fieldwork / debrief	10d	23/09/02	04/10/02	Researchers
17	Fee structure confirmed		03/10/02	03/10/02	Pay & Admin (B)
18	Deadline for last CAPI spec changes		08/10/02	08/10/02	Researchers
19	ARF/admin block confirmed/ARF to print		10/10/02	10/10/02	Researchers
20	Showcards finalised and sent to print		10/10/02	10/10/02	Researchers
21	Finished CAPI program ready for de-bugging		15/10/02	15/10/02	Projects Computing
22	Program hand-over to Brentwood for de-bugging		15/10/02	15/10/02	Projects Computing
23	De-bugging/testing of finished program (L&B)	5d	15/10/02	21/10/02	Projects Computing
24	**Sample Preparation & Despatch**	**7d**	**11/10/02**	**22/10/02**	
25	Opt Out Mailing		11/10/02	11/10/02	
26	Addresses selected, file to programmer		14/10/02	14/10/02	Researchers
27	Label/mailmerge files available		17/10/02	17/10/02	Projects Computing
28	ARF supplies received in Brentwood		17/10/02	17/10/02	Researchers
29	Remaining print supplies in Brentwood		17/10/02	17/10/02	Project Teams
30	Prepare ARFs / sample cover sheet	3d	17/10/02	21/10/02	Project Teams
31	Prepare advance letters	3d	17/10/02	21/10/02	Project Teams
32	Despatch to briefing		22/10/02	22/10/02	Project Teams
33	**Control System Set-up**	**10d**	**23/10/02**	**05/11/02**	
34	Control system set up		23/10/02	23/10/02	Projects Computing
35	Transmission test of de-bugged program	1d	23/10/02	23/10/02	Project Teams
36	Loading of allocations	10d	23/10/02	05/11/02	Project Teams
37	Reports set up and running		25/10/02	25/10/02	Projects Computing
38	**Main Fieldwork**	**31d**	**24/10/02**	**05/12/02**	
39	Briefing period	10d	24/10/02	06/11/02	Researchers
40	Fieldwork period	30d	25/10/02	05/12/02	Interviewers
41	**Post-fieldwork Processing**	**40d**	**05/11/02**	**01/12/02**	
42	Listings available		05/11/02	05/11/02	Projects Computing
43	First coding spec to programmer		19/11/02	19/11/02	Researchers
44	Code frames available / Final coding spec		19/11/02	19/11/02	Researchers
45	Fact sheet complete & available		03/12/02	03/12/02	Projects Computing
46	Coding program available and tested		03/12/02	03/12/02	Projects Computing
47	In Office Edit / Coding (Summary Task)	17d	04/12/02	27/12/02	Project Teams
48	Clean data available		31/12/02	01/12/02	Project Teams



Symbols
Critical
Milestone
Planned by Powerproject Teamplan

This first breakthrough in CAPI had a dramatic effect on NatCen's modus operandi. We no longer do PAPI surveys and even postal surveys take up a relatively small part of our income. The impact of CAPI on the make-up of employment within NatCen has been dramatic. As Table 2 shows the ratio of clerical staff to research staff has halved over the decade that we have moved from PAPI to CAPI. We now have a ratio of roughly one researcher to one clerical staff whereas at the beginning of the nineties it was one to two, a dramatic increase in productivity.

Table 2: Trends in Operational and Research Staff

	No. of Operational Staff	No. of Quant Researchers	Operational Staff per Researcher
July 1992	37	19	1.9
July 1993	35	24	1.5
July 1994	44	28	1.6
July 1995	48	36	1.3
July 1996	47	36	1.3
July 1997	45	34	1.3
July 1998	47	41	1.1
July 1999	52	45	1.2
July 2000	59	45	1.3
July 2001	61	48	1.3
July 2002	65	62	1.0
July 2003 estimated	70	70	1.0

Despite this we still have some way to go in weaning our survey managers and interviewers away from paper. Interviewers are still supplied with paper details of their individual assignments and return paper copies of the outcomes of each assignment, which are kept for record purposes. This reflects a mixture of problems in electronic communication with freelance interviewers; technophobia among older staff and interviewers together with a genuine concern for the quality of the data; the clunkiness and weight of our laptops in the context of the initial contact with participants; the crowdedness of our management agenda and the cost of change. We have set ourselves the target of significantly reducing this sea of paper in the next two years.

In the 90s also NatCen began to build up its CATI facilities, impressed by the ubiquity of CATI for similar purposes in the USA, for example. This remains on a much smaller scale, however. Compared with our annual capacity of about 100,000 interviewer-days in CAPI we estimate our CATI capacity at about 80,000 interviewer hours and we shall only reach that in the course of next year. Part of the reason for this difference in capacity has been the fact that we only use random samples for our surveys and so, in the case of telephone surveys, we have only used samples whose telephone numbers we had access to. We have not been convinced until recently that Random Digit Dialling (RDD) was practicable in this country, in the sense of delivering sufficiently unbiased samples to allow robust conclusions to be drawn from the data.

This attitude is now beginning to change, thanks to some work by my colleague Gerry Nicolaas of NatCen's Survey Methods Unit and Peter Lynn of Essex University and former head of the Unit. In a recently published paper (Nicolaas and Lynn 2002) they looked carefully at our experience using RDD in two recent studies of elections in Wales and London. Although it had been the case that the use of telephone surveys in social research had been hampered by this inability to select a representative and unbiased probability sample of the general population, recent changes in the UK telephone numbering system and advances in telephony software seemed to open up the way to using RDD.

Whilst in the 80s as many as 20% of the population did not have a phone, by the end of the 90s this had fallen to 4%, although this was still disproportionately among the deprived members of the population. However, the proportion of ex-directory numbers is very high and spatially skewed, with

over 56% of the residential lines in London ex-directory and the demographics of those who are ex-directory is also highly skewed towards smaller and poorer households. This renders the use of telephone directories inadequate for a general population survey.

The ability of advanced telephony systems to identify and screen out over 95% of ineligible numbers makes RDD considerably more attractive in a UK context. In the Welsh case this meant that the proportion of private households among the remaining numbers was about 60%, with the remaining 40% a combination of business numbers and non-working numbers. So Nicolaas and Lynn concluded that, in terms of obtaining representative samples for general population surveys RDD is now a feasible survey method in the UK. "Sample design and selection are easy and efficient. Fieldwork is relatively efficient but could be made more so with the help of further research. However, given the dynamic nature of the telephone market [in particular the growth of mobile-only households], telephone coverage will have to be monitored regularly."

For Nicolaas and Lynn the major challenge in the UK was that posed by low response rates. In both the Welsh and London surveys response rates were less than 40% and response rates of this kind would not be considered adequate by most public commissioners of social surveys. The challenge for those wanting to use RDD for rigorous social research is how to increase these response rates. In all probability, as I suggest later, this has more to do with the human aspects of research than with the technology.

Some of the research we do involves asking people about intimate details of their life such as sex and crime, where respondents are more likely to co-operate if they can use a self-completion methodology. One of NatCen's better known surveys is the National Survey of Sexual Attitudes and Lifestyles. This was first carried out in 1990 to inform HIV/AIDS projections and sexual health strategy and used a PAPI methodology. As part of the feasibility work for the second study to be carried out in 1999/2000 my colleague Bob Erens and his collaborators (Erens et al. AIDS 2001) carried out a methodological experiment to test whether the use of Computer Assisted Self Interviewing (CASI) was acceptable and feasible for the general population, and could obtain data of similar or improved quality to pen-and-paper methods. The effect of CASI was measured by comparing rates of reported behaviours in respondents randomly assigned to CASI and PAPI.

The results reported by Erens et al. indicated that CASI was highly acceptable, that there were no differences in unit response rate between the two methods but there were substantial improvements in data quality with much lower rates of item non-response. Qualitative research carried out to aid the design of the study showed that even those with no computer experience found CASI easy to use, appreciated the chance to correct mistakes and felt that the method enhanced confidentiality. The lower rates of item non-response of CASI reflected the programmed routing of questions. In PAPI, respondents needed to follow written instructions, which they may have misinterpreted or misread, but in CASI once a question was answered the programme moved to the next relevant question. Nor was there any evidence of that CASI, in general, affected reporting rates according to the sensitivity of the question asked. CASI methods were, therefore, used in the 1999/2000 survey which interviewed 12000 people in Britain between the ages of 16 and 44, with a urine sample being taken from half of them to test for the prevalence of Chlamydia.

CASI is now a well established technique in social research but this year NatCen has initiated the first large-scale use in Britain of Audio CASI in the Criminal Justice Survey it is carrying out for the Home Office (in collaboration with BMRB). This survey is asking 10000 adults and young people from the age of 10 if they have taken part in any criminal activity of various types and what contact they have

had with the criminal justice system as a result. They are also being asked about their use of illegal drugs and alcohol, their mental health, experiences in childhood and other lifestyle questions.

To ensure privacy and encourage honest responses pre-recorded questions and answers are read out to respondents through headphones attached to the laptop, so that interviewers do not even know which questions have been answered, let alone what the respondents answers were. Clearly this approach is particularly useful among people whose reading ability is poor. Researchers believe that the Audio CASI approach has contributed to achieving response rates (about 75% for the core sample) which are about 5%-10% better than was expected based on results for the British Crime Survey which is a survey of victimisation.

What about the further reaches of new survey technology such as Web surveys and SMS text messaging? So far I have to say that we remain underwhelmed by their potential for our sort of research. Because of the importance to our customers and to our research methodology of representativeness, the highly skewed current incidence of web-use and text-messaging suggests to us that it will be some time, if ever, before these methodologies can be used in most of our surveys, where interview lengths tend to be long and the content complex. In our recent Youth Cohort Survey we offered respondents a web-based alternative, thinking that this population at least might want to take advantage of it. In fact in the first year only 1% of respondents took up our offer and this has risen to only 10% since, even in this young population. The prospect of carrying out one of our two and a half hour English Longitudinal Study of Ageing survey interviews using text messaging makes me literally feel weak !

When faced with the hype surrounding technological trends I remember my first job with Shell Oil in 1970 where I had responsibility for forecasting transport fuels in the UK. One of my projects was to investigate the consequences for petrol sales of the rapid spread of telephones among the population which raised fears that within a few years substantial proportions of the population would be phoning their work in from home and not needing to commute to work and buy petrol ! Needless to say I suspect that, among all the geo-political and environmental, not to mention technological, effects which have occurred over the past 35 years, it must be quite difficult to isolate this particular trend, and whatever the difficulties of the world's oil companies this is not likely to rank high among their risk factors !

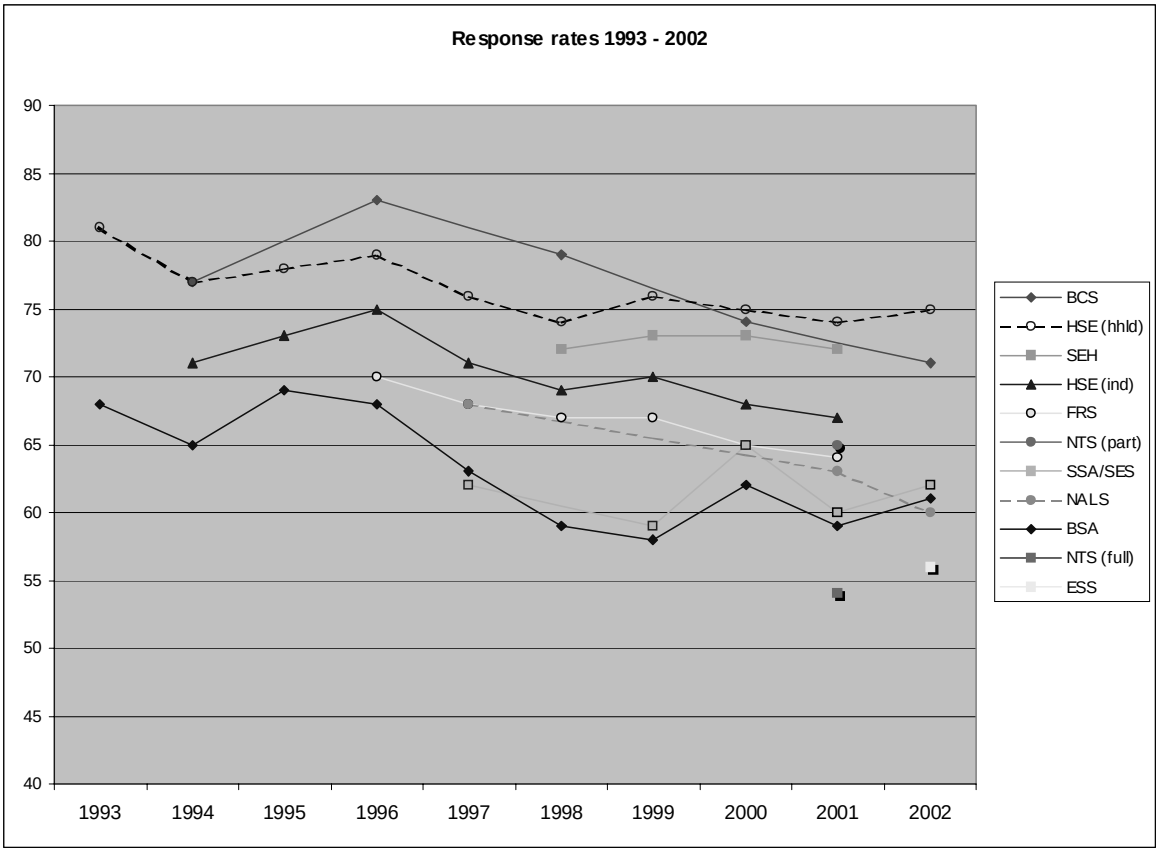
4. The Future for Social Research

Computer Assisted Interviewing has undoubtedly transformed the way we work. Not least it has made it much easier for us to monitor the work of our fieldforce and improvements in ICT should allow us to improve significantly our monitoring of travel time and costs and the time spent interviewing. All of this should allow us to be much more efficient in the use of our fieldforce, which is now the key constraint on the growth of our business. We have been somewhat surprised, however, despite a lot of the hype, how little penetration there has been by CAPI in much of the conventional market research sector, outside a few large firms. Certainly when we recently sought to supplement our fieldforce by sub-contracting some of our work we could not find any conventional market research companies who were able to provide a pool of interviewers trained to carry out lengthy CAPI interviews at pre-selected addresses.

Despite the great boon of CAI, I do not at present see a technological white knight on the horizon to help us with the continuing problems of carrying out rigorous social research in this country – indeed

many of the difficulties are probably never going to be susceptible to technology. The advantages of probability sampling in delivering representativeness depend on high response rates but it is clear that the last decade has seen a secular fall in the level of response rates not just for NatCen but for all providers of high-quality social research. Chart 3 shows the trends in response rates for some major UK surveys through the 90s. It is clear from this that we have a major problem on our hands if we are to avoid probability sampling simply becoming an expensive form of quota sampling.

Chart 3: Trends in Response Rates



Much of this decline is due to socio-demographics, with smaller households and more working women making contact more difficult. Fear of crime and the weakening of the social cement may also play a role, along with a rise in the male gender balance of interviewers. But no trends are inevitable and it is up to us to counter this one by better training and selection of interviewers, for example. We already know that there are significant differences between our interviewers in the success with which they encourage respondents to participate. We need to understand these differences better.

Our interviewers tell us that one of the three things they most like about working for us is the interesting content of the work. Selecting interviewers for this trait is essential in ensuring a high quality of data for our customers and for social research in general. Having our interviewers understand and identify with the research being carried out plays a vital part in ensuring that the data we obtain on incomes, life-experiences and so on, is accurate. It will be accurate because they will take the time to empathise with respondents and to build up relationships of trust which can be vital. Much of our work, for example, consists of longitudinal studies or panels. Sympathetic interviewers are vital in persuading respondents to take part once, twice or even three times more. Technology can play a small role here but we need more than anything else to think and act smarter.

Collecting better data is not the only issue in social research. One of my fears is of a general "dumbing-down" in the quality of social research. I was a civil servant for 25 years and, at various times managed branches that commissioned social research in various ministries. Since 1997 we have seen a big rise in the proclaimed desire to implement "evidenced-based policy" and a commitment to research and evaluation which is very laudable. Alongside it, however, I detect a growing impatience with the long lead times necessary to carry out research that can deliver credible answers. Of course research that is not timely is worthless but the converse is not true – research that is timely can be worthless too if it is carried out in such a way that it does not answer the questions posed or uses evidence which is not credible. Indeed sometimes it is better to know nothing than to know things which are actually wrong !

5. Conclusions

Technology has transformed the way we carry out rigorous social research in the UK. CAPI and CATI in particular have enabled us to extend the boundaries of feasibility and to increase productivity. But many of the underlying problems of social research such as declining response rates, ensuring accuracy of data and maintaining the demand for high-quality research among customers are likely to be less susceptible to technological fixes. Instead they demand better management, better training, selection and inspiration of interviewers and a continuing missionary effort among the politicians and commissioners of social research.

References

Erens B et al. "Effect of Computer-assisted self-interviews on reporting of sexual HIV risk behaviours in a general population sample: a methodological experiment" **AIDS** 2001, Vol 15,No 1 pp111-115

Nicolaas G & Lynn P "Random-digit dialling in the UK: viability revisited " **J. R. Statist. Soc. A** (2002) **165** *Part 2* pp. 297-316

About the Author

Norman Glass is Chief Executive of the National Centre for Social Research (NatCen), the UK's largest independent social research organisation. Until March 2001 he was a Deputy Director in the UK Treasury's International Finance directorate and was previously a Deputy Director in the Treasury's Public Services Directorate where he was Chief Micro-Economist. He had a primary role while at the Treasury in developing the Sure Start programme and is Chair of Croydon Sure Start and High Scope UK as well as being a Board member of the Countryside Agency. He was previously Chief Economist in the Department of Social Security and the Department of Environment. He was the Chairman of the Economic Policy Committee of the European Union from 1999 to 2001.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Official Statistics and Their Place in Democracy

David Pullinger

Abstract

With the establishment of National Statistics in 2000, the long-entrenched view that the role of national statistics was to inform Ministers of State gave way to the democratic assumption that national statistics provide the means by which citizens hold their government to account. This paper discusses the challeges of making this information available and the integrated apprroach adopted by the Office of National Statistics to meet them.

1. Introduction

Information is the oxygen of rule by the people. So believed the Prime Minister in establishing National Statistics, launched in June 2000:

> *'In the consultation document "Statistics: A Matter of Trust", the Government emphasised its intention to seek a new relationship with citizens based on openness and trust.*
>
> *I believe that having access to official statistics which we can all trust is essential in any healthy society. Statistics encourage debate, inform decision making both inside and outside government and allow people to judge whether the Government is delivering on it promises. For official statistics to play that key role effectively in a democracy, we need to have confidence in the figures themselves.' (1999)*

He echoed the words of another European leader, President Jospin of France, in 1989:

> *'The right to information has become one of the fundamental rights of the twentieth century citizen. In a society where information and the media play a considerable part, your [professional statisticians'] action helps safeguard a fundamental human liberty. The work methods you use are complex, the data you deal with difficult to evaluate. An effort to explain [to the public] is necessary. This effort is required by democracy. All citizens must be in a position where they can understand assess the polices followed by governments.'*

Their kind of democracy assumes that citizens get the information able to hold governments to account on their policies. This view in the UK overturned decades of pursuing a line that official statistics were there to inform Ministers of State (The Rayner Report, 1980).

But how do you inform the citizen? How do you enable them to get hold of data that covers the whole of living, from the state of the national economy through to the details of local environment, from the childbirth rate to assessing the impact of Internet, from the number of hatmakers to the overall steel production in the UK?

The Office for National Statistics decided on an integrated approach of making statistical information available and seeking means of alerting citizens to know about it and that it is relevant to them, easily accessible and trustworthy.

2. From niche print to web broadcasting

First the decision was make in 2001 to exploit the new electronic communication and put all statistical information onto the Internet and, in particular, the Web. At that time the only data available on the National Statistics website were in databases and designed for professional use, so a radical change was needed.

Statisticians and communicators now look for the (statistical) story in the data being released and present that upfront. This moves the organisation into a 'news' approach to releasing data and the website now reflects this sense of broadcasting with a design similar to media sites rather than to a print deposit.

Figure 1:

www.statistics.gov.uk, the home page for National Statistics Online, illustrating the news-led approach to statistical communication and the tansparency of corrections and alterations.

3. Topic-based reporting

Not everyone seeks the latest release of statistics - they are interested in a topic and knowing what is available on that topic. This is most likely to be the case for the citizen in contrast to a professional. Consequently we have taken all the statistics produced across government, dividing them into topics and provide a summary with links to datasets and topics of related interest. Each consists of one graph or table and up to 250 words of text.

This collection of topic-based summaries is called 'UK at a glance' and kept up-to-date as each statistical release appears. In traditional parlance, it could be described as a yearbook that is updated daily. Anecdotes indicate that not only citizens find it helpful. Professionals find it invaluable for guiding them to the articles and data available on a particular subject, as it is designed as an integral part to help users navigate to information.

Figure 2:

The contents list of UK at a glance

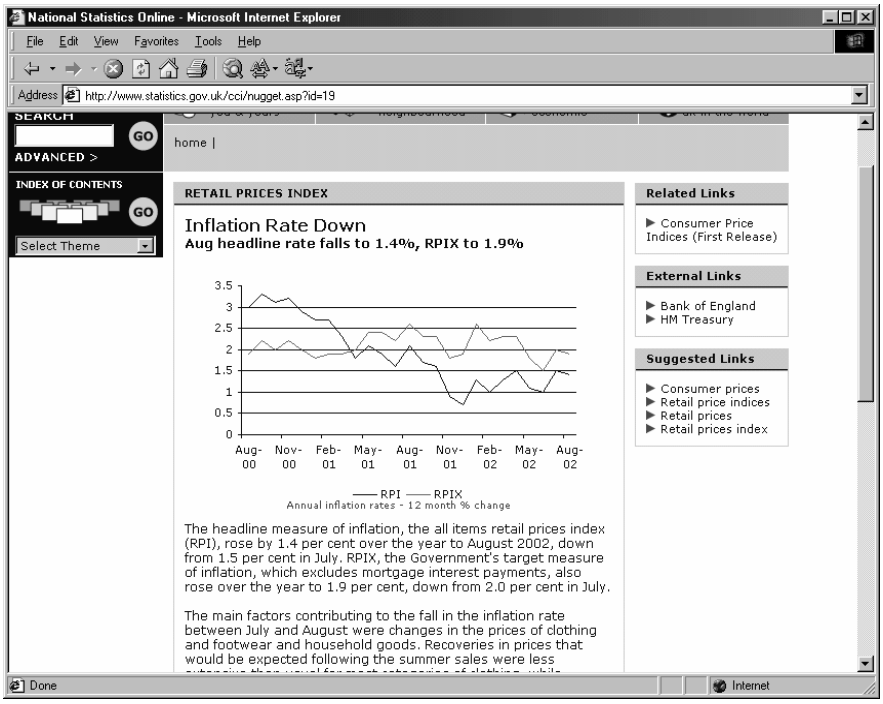


Figure 3:

An example of a topic-based summary. Note the use of the generally understood term inflation as well as the statistically accurate measures

Data can be collected on a particular topic in several different surveys and sources (for example the number of unemployed or single parents). The approach uses the expertise of National Statistics staff to guide readers to the measure considered more reliable, with explanation as to why that is so, rather than leaving the reader to try and sort it out.

4. Geography-based reporting

Most citizens probably relate most to what is happening to them locally - crime, pollution, schools, healthcare, house prices and so on. How can they find out? Such statistics have appeared deep within structured databases and are hard to find. By relating data to geography, one can drill down to a relevant area and find information, or ask questions by 'show me where....'. Simply by putting the

information in the right place (!), referenced by its location, as an exercise in information management and statistical infrastructure unlocks the richness of data previously hidden.

In our focus groups on usability of the Website, citizen group members often declare that they have no interest in statistics at all, until the local dimension is explained to them, and they all say that they do indeed find that invaluable - particularly when choosing where to live if they are moving.

But in the same way that the Prime Minister declared the value of statistics for holding national governments to account, we can expect that a comprehensive small area statistical service could do the same at the local level - holding local government to account for their prioritisation and policies.

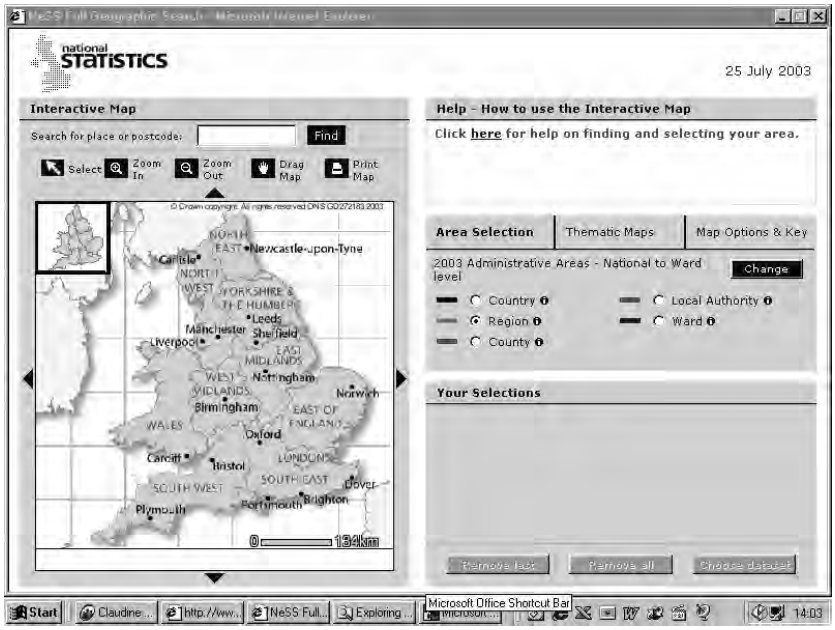


Figure 4:

The entry point for drill-down into geography-based data

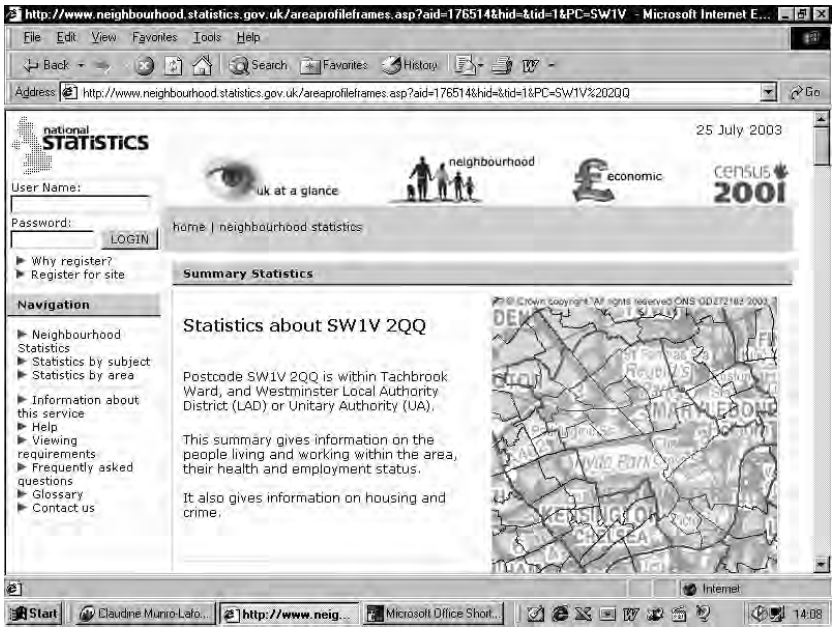


Figure 5:

Information about the Ward in which the London Office for National Statistics resides

5. Raising awareness

But all this development would be wasted if people didn't know about it, if citizens didn't know that the organisation publishes information that might be relevant to their living and able to be trusted.

Consequently the changes to web-centric publishing is matched by seeking greater coverage through broadcast and popular media. Web development and mass awareness are kept in step.

In many countries this is not a problem, the national statistical institutions can issue a piece of statistics and expect it to appear in the press. In the UK one has to sell the idea to the media in order to convince them that their readers would find it of interest, not just relevance. Over the past year we are getting a lot smarter at introducing statistics to citizens as interesting and important to their lives.

The release of Census 2001 results, for example, amounted to the equivalent of over £17m worth of TV and media coverage that not only communicated the value of their participation but also introduced many to the data resources available. The Website now gets over a third of a million visitors each month, the majority of whom are known not to be professional users because they are searching for it on the major Web search engines such as Google and Yahoo.

Figure 6:

Example of the extensive media coverage gained by planned events to garner interest in statistics - although the headline was not promoted.

Moreover in a strategic partnership programme statistics are placed into many different services, so data become accessible in places where people are already looking - for example in the local information services, such as UpMyStreet and the BBC equivalent.

And because not everyone has the Web, ONS runs a centralized Customer Contact Centre handling over 20,000 enquiries per year through phone and email. In particular they support those who would otherwise be disadvantaged by a web-centric approach - those who have difficulty understanding written English or through some disability find the medium difficult to use.

ONS has introduced an integrated communication strategy encompassing media awareness, web publishing, phone and email support. To accomplish this, the National Statistician took the courageous and unusual step of appointing communication professionals and experts to key posts to help statisticians with translating information into forms suitable for citizens and media. If a statistical service, independent of direct government influence, does not proffer the story then interpretations might be invented without the depth of knowledge and background (with some notable exceptions

among newspaper journalists). By providing clear, relevant and accessible information, statisticians are not only managers of data-delivery processes, but also have a role in supporting democracy.

6. Layered statistical communication

All these are part of a much broader change in statistical reporting - towards one that has a new layer of accessibility in plain English at the top of all written content; a new layer that can be extracted and used as part of informing people about the content following.

In technical terms, this new layer is constructed as Discovery Metadata - the records on which search engines operate, indexes and lists are created, and readers can have explained to them the context and relevance of the report or data.

A systematic layering in this way will also apply to reporting primary surveys - such as RPI or General Household Survey. We don't yet do this, so there are no simple descriptions of all the surveys we do and how we carry them out, or the questions we ask in surveys.

7. From improving communication to transforming statistics

These developments in National Statistics are essential. Official statistics would appear increasingly irrelevant to the needs of the nation if no attempt were made to ensure a more populist communication. Almost all major organisations have invested in creating information environments that are more populist and relevant to the needs of their customers - whether it is buying airline tickets online or declaring their environmental policies and other social responsibility programmes. Current culture expects good communication that is accessible to the general user. The information environment in which official statistics are released has changed.

But it is not only the information environment that has moved but also the statistical one. Statistics are now discussed more widely from, for example, the impact of an ageing population through to the five Euro tests. These issues may be the result of statistical research or be a question of the day to which statistical evidence is sought. Or the challenge might come from the recognition that in many cases the statistics historically gathered have been used more to measure the operations of the state agencies dealing with the matter rather than matter itself.

Government itself makes more strategic use of statistics in its performance measures and evidence-based decision-making. Reliable and understandable statistics are forming a greater base for policy-making than hitherto, even if not impinging directly on policy creation. This is particularly so in the context of making already stretched resources go even further. Just as if the ways of communicating statistics stayed the same, the changing environment would make them look increasingly out-of-date and irrelevant, so is it true for statistics themselves.

The development of a cross-Government small area statistics service is a good example of this. The Neighbourhood Statistics Service was initiated by the difficulty of finding information relating to a particular area and the need to focus investment into areas of greatest need and monitor the results over time.

Both raise expectations. More understanding of official statistics elicit demands of even more - and often areas that are less tractable to easy measurement. The value of many UK measures is their longevity, the on-going collection of same or similar measures over very long periods. However that

also means the steady accretion of measures that might no longer answer the kinds of questions being asked and reflect the changing national and international needs. Statistical offices constantly have to develop new measures which not only meet international expectations but also seek to measure social and economic phenomena that can be somewhat vague and intangible.

In the UK, this includes pressures for statistics that describe, for example, the effect of the new economy; the impact of globalisation; pension contribution statistics; information on the distribution of wealth and income, particularly in the area of low pay; and the disaggregation of statistics to support regional and devolved government.

Indeed there are six main areas of change:

Source of challenge
STAKEHOLDERS ARE EXPANDING IN DIVERSITY AND TYPE:
New stakeholders arise from citizens, globalisation, world treaties and organisations, multinational organisations and services, global property rights
THE QUESTIONS ARE CHANGING:
New issues arise from ageing populations, the information age, environmental change, regional governments, smaller government
WHAT IS OBSERVED IS CHANGING:
Deregulation, privatisation, the services economy, new household forms, forms of business organisation, monitoring of transactions, change the world we measure. Customs attitudes, language and values change.
THERE ARE MORE ALTERNATIVE INFORMATION FORMS:
There is a changing balance between official statistics, scientific research, market research, and commercial statistics. There is more chance that other forms of information exist, internationally, commercially, other processes.
THERE ARE NEW STATISTICAL MODELS and PROCESSES:
Access to administrative information from government and commercial processes, information management and technology, and privacy change statistical processes.
THE INFORMATION ENVIRONMENT IS CHANGING
Are the results meaningful, accessible and applicable. The internet changes opportunities to get access to information.

Figure 7: Six sources of challenge and change for statistics

Moreover the ready availability and greater use of statistics by those untrained raises questions about whether they are used for purposes for which they were never intended, thus undermining perceived quality. It is not only statistics that need to be available but also metadata about statistics and ways of presenting them that indicates appropriate use.

We need a new responsiveness, flexibility and depth in the statistical infrastructure.

8. Increasing trustworthiness

Having an independent department lead official statistics is only one step towards public trust. There is also need to codify practice to ensure that there cannot be ministerial interference in their content or

date of release. This is what lay behind the introduction of the National Statistics Code of Practice and its various protocols.

Other main factors include a willingness to convey an authoritative and objective understanding of what is happening to the UK, its people, business, communities and environment - that we have sought to do through the integrated communication of web broadcasting and mass awareness. And an openness and transparency not only to process but when things go wrong. When data is changed or (in extreme situations only) withdrawn for a period, a message is given on the front page of the website.

We can afford few risks on quality, needing to improve the reliability of our statistical systems, the lack of which has necessitated the visible revisions that has at times made UK statistics vulnerable to criticism, and ensuring the skill levels of staff continually rise.

9. Statistical modernisation

Many of the developments required lie in the area of data collection and storage. Without common classifications, for example, one cannot easily accomplish cross-cutting data. Integration of collection and data storage enables flexible responsive analyses. Many of the hard public policy questions necessitate understanding drawn from disparate sources - for example in the areas of family, culture, ethnicity and well-being and work. Integration as a planned characteristic of the statistical system means that ONS will have the capability to answer questions to issues that were unknown when at the time when the surveys were originated.

The Government has invested £75m to modernise ONS statistical infrastructure:

- Putting in place common methods and practices
- Introducing fewer and more contemporary tools and software
- Creating a single common data warehouse for all ONS statistics
- Re-engineering key statistical processes
- Collecting more data electronically and expanding use of administrative data
- Disseminating statistics electronically in a highly relevant and accessible form
- Shifting staff skills from routing checking and assurance to higher value added activities.

This programme of modernisation begins in 2003 and is due for completion in 2006, replacing the patchwork of statistical processes, systems, methodologies and software inherited from the component organisations that made up Office for National Statistics when formed.

10. Conclusion

In this presentation I have focused on the citizen, because for the UK this is a dramatic change in approach from the previous one that emphasised the role of official statistics as giving information to government Ministers. They are but one stakeholder among many. However, as the National Statistician has noted (2002), perhaps one of the most important aspects of the new 2000 framework for National Statistics is the recognition that Ministers are most likely to trust a system that the public can see they have good reason to trust. In other words citizens' access to and trust in official statistics are likely to lend support to evidence-based policy-making rather than simply being doctrine-led.

Focus on the citizen does nevertheless reveal the need to modernise the official statistics infrastructure in order to be more responsive to the questions of the day. Much of this is in the harmonisation of

collection, use of sources and integration of data. ONS is investing in this modernisation and will have a 21st Century statistical system in 2006 that will support the ideals set out for providing trust in data essential to a healthy democratic society.

About the Author

Dr David Pullinger is the Deputy Director Communication in the Office for National Statistics, UK and can be contacted at david.pullinger@ons.gov.uk.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Defending the Defensible. Protecting the Status of Survey Research

Peter Mouncey

Abstract

This paper argues that the survey research industry has a proud, and enviable, history based on technical competence and adherence to self imposed ethical principles. This has led to the industry enjoying more freedom than other sectors within consumer protection legislation – despite two challenges by the data protection regulator in the UK. Technological developments provide the industry with new opportunities, but these self-same technologies also lie directly, or indirectly, at the heart of these challenges, and this is likely to continue into the future. In parallel, there is a growing awareness within society about the practices used to collect personal data and the industry's self regulation procedures, leading to a growth in complaints to the bodies responsible for policing the industry. The paper describes firstly the need for ensuring that the processes used to enforce standards are totally transparent, objective and professionally managed and secondly, the continuing action being taken to defend the current self regulated status of the industry within the UK, and on the wider European front, to ensure that the industry remains in control of its destiny.

1. Survey research: a brand with a tarnished image?

> *'I'm a 25 year old florist. Well that's what I told the last survey I completed. So how reliable can they be?' – Joan Bakewell, Guardian, 28-02-03*

The ground covered by Bakewell gets right to the heart of one of my main theme – the impact of image on the self regulated status of an industry claiming to have its foundations built firmly in the realms of pure science. She goes on to say:

> *'But surveys themselves need checking out. They are not value-neutral, laboratory-style findings. They are often done in the street – in bad weather – by people with clipboards, who would prefer to be doing something else, with passers-by who do their best to avoid eye contact but relent before the supplicant's 'It will only take a minute'. So why is this far from exact method used as reliable evidence for claims that can frighten us, threaten us and cajole us – but rarely reassures us?'*

This brief paragraph is bursting with issues and contains several uncomfortable echoes back to the findings from the major study on non response undertaken by the Research Development Foundation in the mid 1990's under my chairmanship. The 'minor urban pest', as one respondent in the project termed on-street interviewers, being one example that has firmly stuck in my mind.

This is not just an attack on commercial market research, Bakewell goes on to implicate all, questioning the use of survey research methods and findings by charities – Cancer Research UK, Sport England - and the dramatic, even frightening, tone in which they often present research findings – to encourage a donation, engender a change of attitude or lifestyle etc.

One final quote from Bakewell's article:

> *Surveys are undertaken for a specific purpose by institutions who consider it money well spent to make a bit of a splash. Surveys, if their findings are startling enough, are sure to make headlines and create news. The more startling, the better. They battle to catch the public imagination, promote future articles and discussion on afternoon television. They seek to fill space – column inches – rich with speculation. Over the years they have helped fuel the popular addiction to health and diets that is surely unhealthy in itself. I'm well aware that they can be used to promote television programmes: remember '72% of the population believe in God' and '70% of women give up sex at 50'.*

And isn't that what we often seek – recognition, influence, debate, and most of all some action by those in positions of power and responsibility.

Articles such as this do nothing positive for our image, especially when from such a well respected source.

2. A new landscape

There's nothing new or remarkable about this attack, so why draw attention to it in the latest round? The reasons are several, varied and inter-related:

- A more knowledgeable, cynical and litigious population;
- The growing impact of technologies;
- Increasing world-wide focus on human rights issues within legislation, especially in terms of personal data and generally protecting the rights of the consumer – the role of self regulation versus the law to impose acceptable standards;
- A growth in the application of 'survey' methods coupled with a seemingly insatiable demand by some sectors for customer data;
- Increasing difficulty in creating a differentiated product;
- A 'drains-up' within the market research industry over the definition and scope of what we do, and how to reflect this in self-regulation measures.

Many of these factors are inter-related. Here's a simple example related to internet technology: traditional market research companies are increasingly using the web for conducting interviews and transmitting data; the advent of the web has encouraged new players to enter the industry who are not subject to the degree of self regulation imposed on the traditional providers by the MRS and ESOMAR codes of conduct; web-sites such as those for the MRS and ESOMAR make it much easier for people with grievances against the industry to 'know their rights' and to instigate investigative action through the bodies who are charged with policing the actions of its members. So, this means that there are those who can be fingered, and those who cannot in terms of a supposed self regulated industry – a situation that perhaps tempts firstly those being fingered to consider cutting the wire of self regulation and crossing the border, and secondly, tempts those in places such as Brussels to consider that maybe control needs to be imposed from outside…Not an attractive scenario for an industry that has fought hard over the past twenty years to have, and keep, favourable concessions within increasingly restrictive legislation compared to a major client sector – marketing.

I recall that when, as an MRS Council member, I served as PSC Chairman in the late 1980's, it was a very bad year if we received a dozen real complaints a year – few enough for the Chairman to act as the first filter. And we didn't seem to need an MRS Codeline to handle queries from members – again the Chairman was expected to deal with many of these. You only have to look at the MRS Annual

Report to see a stark illustration of how the world has changed in recent years. The table shows the levels of complaints received by the Standards department in recent years.

Complaints Received	**2002-2003**	**2001-2002**	**2000-2001**	**1999-2000**
Disciplinary cases				
Upheld	2	2	4	1
Not upheld	21	22	6	2
Ongoing	9	11	7	2
Total	**32**	**35**	**17**	**5**
Non Disciplinary cases				
Complaints (resolved without use of disciplinary procedure)	N/A*	N/A*	29	20
Minor complaints (resolved by Secretariat)	75	72	54	48
Total	**75**	**72**	**83**	**68**
Total complaints	**107**	**107**	**98**	**73**
Carried over from previous period	11	7	2	0

* Now included in the Disciplinary cases figures. (Source: MRS Standards Department)

Note the uplift in the number of cases requiring investigation by the PSC under the Society's disciplinary procedures. Thankfully for the industry, few have been upheld – but that misses the key point, which is that the MRS are receiving many more complaints, and many of these are sophisticated and complex. Many more people out there feel able to question the methodologies, practices and processes of the industry. Regardless of their agenda, any reasonable query has to be taken seriously. These require considerable time, effort and therefore cost, in order to ensure that they are fully investigated. And, the process must be seen by all concerned as thorough, objective, transparent and providing equality. Hence the need to keep these processes under review and ensure they conform to the principles enshrined in European human rights legislation. It also means that the industry is increasingly under the spotlight and this needs to be recognized by us all.

A further interesting statistic is from an analysis of calls to MRS Codeline. The implementation of the 1998 Data Protection Act and the subsequent publicity this received plus advice and guidance issued by the MRS, led to queries about data protection rising from only a handful in 1999-2001 to 90 in 2001/02 (38% of all queries) and 75 (33%) in 2002/03, demonstrating the impact of this legislation on the industry. The next most common issues in the latest year were researching children (25 queries), mystery shopping (15) and respondent confidentiality (13) – the latter being closely linked to data protection.

Finally, calls to the MRS Freephone services, interviewees checking on the authenticity of a research agency, averaged over 16,500 in the years 2000-02, up from 9,000 in the previous year. The latest figures show a fall to12,000 calls, but this should be viewed in the context of agencies introducing their own services, that in some cases are 24/7 and can check out interviews down to precise geographic areas, plus, a separate on-line service set up to reflect the growth in this methodology.

All of these figures might be seen as miniscule for an industry that probably undertakes over 25m. interviews per year in the UK alone – but that really misses the point of self-regulation. For example, the TV and advertising regulatory bodies often decide that serious infringements of their codes have taken place based on only a handful of complaints from a potential audience of many, many millions of consumers. That's what self-regulation is often all about, whether we like it or not – the views of a concerned few leading to responsible action being taken by the industry concerned, sometimes leading to significant long-term change.

3. Self regulation: an enviable heritage

Survey researchers have an enviable reputation for understanding the need for a strong ethical stance – recognizing from the earliest days of an organized industry the fundamental truth that there would only be a viable long-term future if the trust of the general public was gained and maintained. The International Chambers of Commerce Code of Conduct was developed and adopted by ESOMAR in the late 1940's - the MRS having to wait until incorporation in 1953 before it could introduce a similar code that would be binding on all its members, which it swiftly did (1954). These two codes, although similar in key principles, remained separate until 1999, when the MRS finally adopted the ICC Code. The key section B, detailing the Rights of Respondents, contains principles that pre-date European data protection legislation by over 30 years.

This has given the industry a strong bargaining position when looking for concessions within the first and second rounds of data protection legislation – in specific countries such as the UK and Germany in the early 1980's and again in the 1990's as the EU Directive started to take shape. As you all probably recognize, this lobbying activity has achieved a high degree of continuing success. Confidential survey research across Europe is generally treated in national data protection legislation as a recognised form of statistical and historical research, eligible for some exemptions, whereas organizations collecting personal data for direct or database marketing have to adhere to special, much more onerous European wide conditions.

But this concessionary status is not a right, and time has shown that it needs to be continually defended. Not withstanding successful negotiations in the legislative consultation periods, this status has to date been challenged twice – firstly, in the late 1980's, and secondly in the late 1990's. In both cases the basis of the challenge has been the accusation by the regulator that all of the industry, or certain key aspects of it, have become a tool of the direct marketing industry.

This is directly as a result of the developments in technology – in particular the rise of customer databases and associated applications, especially customer relationship management (CRM), increasingly becoming the basis for all marketing.

In the case of the 1984 act, the issue was very straightforward – the advent of 'lifestyle' surveys and customer database enhancement for targeting direct marketing campaigns. Had we sold our skills, experience – and even our souls to the new marketing analysis industry? After all, for example, both sectors have over the intervening years significantly benefited from the then new geodemographics.

The RDF project on respondent co-operation, conducted a few years after this initial challenge was refuted, clearly underlined why the regulator took this view, and why it remains an issue for the industry. In terms of the confidentiality of personal data collected in market research surveys, attitudes showed a very confused 'joe public' and the need to promote the anonymity at the heart of the Code:

- 50% of respondents thought market research was an invasion of privacy;
- 71% expected to receive 'junk mail' directly as a result of taking part in a market research survey;
- 69% stated they would be less likely to participate in market research if their name & address was passed on;
- 45% thought they would be more likely to participate in surveys if their confidentiality was guaranteed;
- 43% stated they would believe an assurance by an interviewer that the survey was confidential and their details would not be passed on to anyone.

It was not really too surprising that the regulators struck again in the run-up to the implementation of the EU data protection directive in the UK. This time the issue was all about the supposed real purpose, or hidden agenda, when conducting customer satisfaction research – was this a disguised form of direct marketing; a basis for 'one to one' marketing to customers? The whole European industry watched this development with intense interest and some trepidation. But, once again we successfully argued the case that genuine, confidential survey research will not lead to the personal details of those interviewed being used for other purposes, thereby protecting the interests of all those working to the principles enshrined within the ICC Code of Conduct.

4. Strengthening the industry's case

This latest challenge led to the development of a whole family of data protection guidance for the industry in the UK:

- The Data Protection Act 1998 and Market Research: Guidance for MRS Members (MRS) + A basic Guide to the Data Protection Act 1998 (MRS);
- Market Research Processes and the Data Protection Act (DPA) 1998 (MRS/BMRA);
- Market Research Processes (Client) and the Data Protection Act (DPA) 1998 (MRS/AURA);
- (Draft) Guidelines for collecting data for mixed or non-market research purposes ('Category 6') (MRS);
- Internet research guidelines – being developed by the MRS based on the ESOMAR version.

Underpinning all these guidelines are the six categories of personal data based feedback from survey projects that were introduced at this time to steer the industry, and, to help us clarify the boundaries with the UK Information Commissioner:

- **Category 1:** feedback only to members of a project team, all of whom have agreed to be bound by the MRS Code;
- **Category 2**: feedback to the data controller where the intended interviewee drawn from a database has either died, or is no longer at the address shown;
- **Category 3**: feedback to data controllers for setting 'do not re-interview markers' on their databases;
- **Category 4**: feedback to the client, separate to the survey data, of any specific complaints or dissatisfactions for investigation/resolution;
- **Category 5**: feedback of personalized survey data to the client for confidential survey research purposes only;

- **Category 6**: projects where either confidential survey research is not the sole purpose for conducting the project, or where the personal data will only be used in an attributable form.

Increasing use of technology poses interesting issues in terms of controlling the use of data within Category 5. Video tapes from group discussions are no longer the cutting edge here. Clients increasingly demand research output in a form that excites, motivates and presents an un-ignorable call to action. They don't want hard data, they want insights and stories. They also want this in a new media format – why not a CD containing real consumers telling it like it is to illustrate key points from a quantified survey? And why not load this onto the company intranet…….? And what about the trend to 'bricolage', where the input from the research survey is only one input, where the boundaries may get really blurred. Hence the development of data protection process guidelines for the client-side.

5. Stretching the industry boundaries

But Category 6 moves the boundaries of what has traditionally been permissible for market researchers. What this guideline is essentially saying is that providing members adhere to the content of this guideline when undertaking data collection projects that are not solely, or anything to do with confidential survey research, you will not be in danger of breaking the MRS Code. The Information Commissioner welcomed this new category as it clarifies the boundaries between traditional survey research and other, legal, data collection activities. Some in the industry remain uneasy about this direction, but it is facing up to an inevitability that has been confronting the industry for nearly quarter of a century as the potential sources of consumer data have steadily widened, facilitated by technology.

6. Legislation can have a positive dimension

The key underlieing principles within the legislation are transparency and informed consent. This is really the application of common sense. Put yourself in the respondent's shoes. As a direct result of taking part in a survey, would a respondent then get confronted with anything that might be considered unexpected. For example, if a person is interviewed in an *ad hoc* survey, could they have expected to be re-contacted at some time in the near future unless permission was gained in the first interview – I think not. If a survey sample is drawn from a customer database, I think that respondents expect to be told the source, if they enquire – especially if there are any errors in the source data.

And, going back to the whole issue of ensuring a self-regulated future, the data protection legislation in the UK model is good for the industry for four key reasons:

- Legal commentators see the UK version of the Directive as providing a good balance between protecting the consumer on the one hand, and recognizing that personal data underpins many aspects of a modern society;
- Confidential survey research enjoys exemptions and escapes restrictions imposed on mainstream marketing – it helps clarify our position and provides a measure of differentiation;
- This legislation gives more weight to the Code. Protecting the rights of the consumer when processing personal data is no longer simply a 'gentlemen's agreement' from a clients' perspective – it is now the law;
- It focuses the industry on improving standards that benefit all stakeholders in the survey research process.

But, despite all this work within the industry, I suspect that a re-run of the RDF questions on confidentiality issues would not show a dramatically different position today than in the mid 1990's –

so the prospect for further investigations remains, especially as we adopt new technologies such as the internet, SMS etc.

7. Simplifying the rules

If self regulation is to work, those covered by the rules need to clearly understand their responsibilities and those wishing to make a complaint should equally expect to quickly see if they have a case. However, the need for more detail and new measures have increased the complexity and volume of advice and guidance provided to members – just look at the relevant pages on the MRS web-site (www.mrs.org.uk). The Professional Standards Committee is well aware of the need to provide members with key up to date information in a simplistic and easily accessible form. The first step will be to group together all the 'musts' – those aspects of the code and supporting guidelines where infringements could lead to disciplinary action – and separate them from a list of the 'shoulds', describing best practice.

8. A legacy of 'old' technology

But let's not neglect an older piece of technology, beloved by direct marketers and market researchers alike – the telephone.

Soon we will have a further piece of data protection legislation in place when the EU directive covering Privacy and Electronic Communications becomes UK law by the deadline of October 30th this year. Not too much to worry us here in terms of the e channel, but in its submission the MRS has requested an opportunity to discuss the industry's use of automated calling systems in telephone research as the directive clarifies the definition of this technology and encourages methods to reduce nuisance calls. But, it would appear that Oftel wants to pursue this issue with more vigour. Earlier this summer some market research agencies were advised that there licences could be under threat unless they take steps to reduce the number of 'silent' and 'short' duration calls generated by automated dialing systems, power dialers, as these are seen as two examples of persistent misuse, from a consumers perspective, of the telecommunications network - practices that tend to reflect the behaviour of the tele-marketing industry.

In the USA, the usage of telecommunications technology by market research has become the basis for a broader challenge by the regulators. Here, industry bodies are embroiled in a big fight to defend the use of tele-research. To quote the Council for Marketing & Opinion Research (CMOR), 'telemarketing [in the USA] has obviously had a crippling effect on research…….'. The fight is on two fronts – firstly, to prevent the legislators subjecting research to the same restrictions as those imposed on telemarketing, through the Telephone Consumer Protection Act (TCPA) rules currently under review by the Federal Communications Commission, covering fixed, mobile and fax telephony channels. Secondly, where market research remains unrestricted, the public think it is, or should be, covered by the same rules and become even more angry with an industry that they see as violating non contact laws – a 'double whammy' in CMOR's view. Just look at their web site (www.cmor.org) if you want to see the full picture. Currently, MRS members are not required by the UK 1998 Act to screen against the Telephone Preference, or other preference services when undertaking confidential survey research, but this position could easily change. Maybe the American consumer has a valid point……...

To help stamp out 'research fraud' in the USA, the FTC and leading bodies, including the US Postal Services and CMOR, have joined forces to create the Consumer Sentinel Program that investigates complaints from the public, which are then held on a database.

The 'new' internet technology poses similar problems of differentiation for the industry – everyone wants to use this media to collect consumer data.

9. Action: European and global initiatives

So, in the UK we have been extremely successful to date in terms of defending our position against the challenges of the regulators. We remain self-regulated and relatively free from restrictions. But increasingly the issues are of course European wide at the very least; global in many cases.

Two examples. First, our data protection legislation in the UK is a national sub-set of a pan EEU directive – this same legislation sends out ripples throughout the world as a key principle provides safeguards to protect the transfer of personal data to those countries without adequate data protection laws; secondly, the MRS Code is a version of the international ICC Code – increasingly this Code has to provide the industry with a set of globally recognized principles, whilst at the same time providing sufficient flexibility to deal with national or regional differences.

An issue in one country can quickly spread, especially through the rapidly enlarging European market. Increasingly, the industry therefore needs to co-ordinate actions at a European or global level. We also need to keep an ever closer watch on the regulators and legislators, if we are to protect our self regulated status. One trend that is certain to accelerate within the EU is consumer protection.

We should see the final shape of the proposed EU directive on unfair commercial practices emerging from the EU Parliament and Council by the end of the year. This may provide some clues on the future role for self regulation in consumer protection – whether the Directive seeks to rely on self regulation for control, as recommended in the Commission's preamble, or whether the Parliament feels that tougher legislation is the answer. But this does stir up the whole issue of self regulation within Europe, and raises some interesting thoughts for the market research industry to ponder.

Firstly, how might the regulators define market or survey research when deciding on the scope of the industry. Whilst the Information Commissioner in the UK has accepted the argument that MRS members control the use of personal data collected in customer satisfaction surveys, they know full well that there are many other companies out there undertaking this type of work that are not regulated in this way.

Secondly, if the current Code is seen as 'best practice', is it considered by the regulators to provide control over enough of the industry to be truely effective? Remember, membership of the MRS, or ESOMAR, is personal based; the Code is currently therefore only binding on each individual personal member of that particular body. However, self regulation within many sectors, such as direct marketing, is managed by the trade body, thereby covering all employees of a member organisation. So, in comparison, the level of potential enforcement within the market research industry is more restricted. Possibly a regulator or legislator could feel that the whole industry cannot therefore be effectively regulated by the MRS or ESOMAR. Faced with new data privacy legislation similar to the EU Directive and recognizing that the legislation specifically covers organisations, the leading Australian market research industry trade and professional bodies have joined forces to try and

overcome this problem by encouraging agencies to sign up to an industry privacy code. Also, principals of AMRO agencies have to sign up to the MRSA Code of Professional Behaviour.

Finally, there is some evidence from the experiences of the advertising industry that the EC is starting to look for a stronger European focus within self regulation than might be provided by codes written from a global perspective.

The increasing need to 'think global' and 'act locally' has led to greater co-operation around the world in identifying issues that are, or could, restrict the activities of the market research industry, and, to lobby more effectively whenever necessary. In Europe we now have the Research Alliance, set up by ESOMAR and EFAMRO in 2001, to protect the interests of the European market research industry. The Public Affairs Committee is responsible for keeping a watching brief on developments across Europe through liaison with national bodies, and the World Federation of Advertisers – the main global lobbying body for the advertising industry. To date, the work of this Committee has focused on:

- Developing a pan European audit of issues that are either currently restricting the activities of the market research industry, or might in future. This covers countries outside the EU, including some of those soon to join. The topics include not only the impact of EU Directives at national level, it includes, for example, legal measures to restrict pre election opinion polling (eg Czech Republic, France), local controls on interviewing in town centres, access to electoral roll data (both in the UK). It is updated quarterly;
- Developing a database of the main issues affecting market research within the national data protection laws across the EU;
- Reviewing the impact of any new EU Directive and advising on action;
- Assessing any need to improve the effectiveness of current self regulation measures;
- Working with ESOMAR to access the level of European focus in the current ICC Code.

In parallel with this work, ESOMAR has the Quo Vadis initiative, reviewing the definition of market research, in the context of industry developments, the needs of clients, and, the current scope and content of the ICC Code.

This is the European part of the Global Legislative Initiative that includes a similar grouping of the leading industry bodies in the USA, Canada and Mexico – creating the potential for exchanging information and tackling issues that have an international dimension.

10. A well respected industry?

Returning to the start point of this paper, I believe that the real value of survey research to society is well understood by decision makers across all sectors of industry, government. A point constantly made over the years whenever regulation threatens. Even the general public recognize that market research delivers benefits to society. In answer to the question 'Does market research serve a useful purpose' in the RDF survey, 78% agreed to some extent with this statement, but the proportion who 'Agreed a lot' at 31% of this total was well below those who 'Agreed a little', 41% - hardly a resounding vote of confidence. The methods developed in other scientific disciplines and traditionally used by survey researchers have been 'borrowed' by others and trivialized in the pursuit of more sales, eye catching headlines – even providing material for populist television game shows. Sometimes in order to engage certain groups in the population we have to adopt approaches that historically would have been unthinkable. But, if we are to successfully defend our turf in the future then we must remain in a position that enables us to be able to defend our methodologies, applications of our work and the ethical stance of the industry. The battles in the USA over telecommunications based interviewing

suggests that as well as taking issue with the legislators we also need to re-consider how we do what we do.

11. Technology and ethics: new challenges ahead?

Technology will continue to provide our industry with new opportunities but the very same technology used by others, or those outside the reach of self regulation, will undoubtedly lead to the possibility of further confrontation in the future with legislators and regulators. Take the new Radio Frequency Identification (RFID) technology, being trialled in the UK by Tesco to help combat shop-lifting. The chip in the packaging triggers a CCTV to film the shopper when they remove the product from the shelves, and again when they pass through the store checkout.

As pointed out in a feature in the Guardian in July, this development has excited those interested in customer loyalty and gaining a deeper understanding of consumer behaviour. Whilst current usage is apparently restricted to packaging and labeling to aid supply chain management and prevent theft, the next logical step is to build the chip into the products themselves and track what happens to the product, and thereby the users, especially when linked to other database information, such as payment and customer profile details. A new generation of 'shopper-watching' is about to unfold and this should be creating a degree of unease within the Information Commissioner's office. And I'm sure that this will soon provide a whole new fertile ground for survey research – a new toy to play with; a new tool in the toolbox. Research suggests that whilst consumers expressed concern they seemed to resign themselves to the inevitability of it. I would argue that this probably because they cannot see all the far reaching implications in terms of how the technology could be applied. What will be the survey research industry's stance on the usage of this technology, and associated ethics? Something to think about now, before the regulator makes the decision for us.

12. Conclusions

What I've tried to argue in this paper is that our industry needs to be vigilant and increasingly pro-active about the threats to the current level of legislative freedom enjoyed by those working in survey, or market research. Our success to-date in defending our position has been based on two factors.

Firstly, the realization by the pioneers within the industry that future success depended on gaining and maintaining the trust of the general public by taking the ethical high-ground, but they did not have the level of competition for the consumers attention – and their data - seen today. Consumers find it increasingly difficult to discern the differences between confidential survey research and direct marketing, across all channels. That is an issue we can do something about. Ethics must remain a key point for differentiation. Secondly, the fact that decision makers in industry, government etc acknowledge the value of survey research – the essential objective 'window on the world' – won through trust in our methodologies and innovation. Therefore, we constantly need to ensure that our methods stand up to scrutiny and that applications do not trivialize our image – the emphasis on quality standards through MRQSA and ISO helps our case, but I've not had the time or space to cover these developments within this paper.

I've pointed out that that the developments in technology are a double-edged sword. They provide the industry with valuable new tools, but the usage of these same tools by others, and by survey researchers themselves, leads to regulators taking a closer look at what we do and asking us to justify

continued special treatment. To some extent this is linked to the success of what we do – others have applied our methods and techniques for other purposes. But the way we use these tools needs to be carefully considered. They make it cheaper and easier to conduct ever more surveys, potentially irritating even more members of the general public in doing so; never mind the quality, feel the width - putting a new spin on the old quant. v qual. debate.

Those who framed the original codes of conduct fully recognized that the public is the industry's key resource. If we respect the consumer, they will respect, and trust, us.

I've argued that legislation can provide benefits to our industry, but that the need to continually demonstrate that self regulation provides effective policing is likely to become increasingly important in an era of consumer protection. Today's consumer is much more aware of what we do and is more willing and better equiped to challenge our methods and practices – these challenges must be matched with transparent and professional complaint handling processes.

Finally, I hope I've demonstrated that the industry is constantly working behind the scenes to defend the interests of the industry and that the degree of global co-operation to achieve this has grown immeasurably in recent years.

So, to return to where I started – Bakewell is correct, surveys do need checking out, constantly – but lets ensure that *we* do the checking first and thereby minimize the future opportunities for journalists and regulators to make an effective case against the industry.

References

Windle, R. & Mouncey, P. Non-response and public perceptions of market research, Market Research Society conference papers, March 1997

Windle, R. Public co-operation in market research, Research Development Foundation/Market Research Society, 1996

Bakewell, J. Guardian G2, 28-02-03

Shabi, R. The card up their sleeve, Guardian Weekend, 19-07-03

Market Research Society Standards Department

About the Author

Peter Mouncey is an elected Fellow of both the Market Research Society (MRS) and the Institute of Direct Marketing (IDM). He has worked in market research and database marketing for over 30 years, is an ex-chairman of the Market Research Society (MRS), and serves on the Society's Professional Standards Committee and the Public Affairs Committee of the (European) Research Alliance. He also worked for the MRS as a consultant in developing the data protection strategies for the UK market research industry in the wake of the EU Directive and the Data Protection Act 1998. In addition to being a Visiting Fellow at Cranfield University School of Management, the author is also Director of Research at the Institute of Direct Marketing.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Data Processing. Facing the Challenge in a Commercial Market Research Agency

Garj Basi

Abstract

Data Processing in commercial Market Research agencies is facing major change. This paper explores what forces are driving this change and how the world of survey processing will need to adapt and evolve over the coming years. These forces of change can be broadly split into two categories; client driven changes and technology driven advances. The paper discusses the plans we have to embrace this change at BMRB.

Keywords

Data Processing, organisational Management, change management, career development, Dimensions.

1. Introduction

In our fast moving world some things appear not to change at all. I have been responsible for the DP function at BMRB for nearly ten years and the task is in essence the same as it was nearly a decade ago. Many facets of society have undergone a transformation but DP is essentially the same task. The Job Description for DP execs from ten years ago is still virtually identical. Many of the tools we use are the same as they were 10 to 12 years ago (Quantum, Bellview, Quancept). The business dynamics within DP have also been the same for even longer. We have battled to recruit, train and retain enough people for the DP industry. This has been true for me at BMRB and for many of my colleagues within the industry.

I believe over the next three years there will be significant changes that will transform the Data Processing industry. Data Processing functions in commercial MR agencies are about to go through a dramatic and structural change.

First I will discuss the forces that are building up which will be the agents of change. Then I will discuss the impact of these pressures and forces and the actions we have planned at BMRB to cope with this new world.

2. Forces of Change - Facing the Challenge

These agents of change broadly split into two categories, first there are customer needs in the context of a tough economic climate and next there are technological factors. There are also interactions between these categories; technology enables business to advance and improve its products.

Globalisation for example became an obvious business strategy when the Internet and communications technology had matured sufficiently.

First, let's look at the demands from our customers.

We are all facing **economic pressures**, which means we face competition to deliver better services. We have to find a way of achieving fast delivery, excellent quality of service, and yet maintain a competitive price. In addition we have to cope with the complications and issues raised by Globalisation. To add to this Shareholders demand growth which means acquisitions, mergers and expansion. We also have to acknowledge that some of our clients use of data is becoming more sophisticated. Finally legislative changes cannot be ignored.

I will now look at these in more detail.

We are in a **price sensitive** industry where our products and services are often not fundamental to the operation of our clients core business so in difficult times we get squeezed further or worse contracts get cancelled. Of course (thankfully!) this is not true for all clients.

Although the DP component of a survey may only be 10% of the job in terms of value – nevertheless pressure remains intense to keep this cost as low as possible. At the same time we need to **deliver error free survey results fast, and to schedule.**

Globalisation now adds another level of complication

I have read some reports that Globalisation is becoming less popular as a concept in the business world. Apparently terrorism, war and disease in the shape of SARS is making CEO's rethink their global strategies of the last century.

I will ignore this for now, as I believe Globalisation is ongoing as far as our industry is concerned.

We have survey collection processes that continue outside normal working hours. Web services and Cati units can collect data 24 hours a day. Our Clients are also spread around the world and want access to their data round the clock.

Obviously the associated helplines and technical support services also need to be available around the clock.

Some clients' information needs are becoming more **sophisticated**. For some of our clients at BMRB data from our surveys is fed directly into other systems. Examples of systems where survey data interfaces with client systems include:

- CRM (client relationship management),
- Remuneration and bonus payments based on customer satisfaction,
- Evidence based policy making.

For the Survey industry these sophisticated clients are great news. We need to nurture them and ensure we build intelligent systems that integrate our survey results with their needs and the interface is not a complicated error prone one. To define and build the right solutions we need to ensure the appropriate level of technical dialogue takes place between supplier agency and client.

Before we look at the technology side we cannot ignore the new legislation and Quality standards that we must embrace. Data Protection, Maternity Leave, Paternity Leave, Flexible working, Home Working are some examples. We need to embrace the spirit of the law by understanding the value of the legislation rather than reluctantly and begrudgingly accepting each one of these.

What this means is that the survey processing or DP function must have the following properties in order to satisfy the market conditions:

- Flexible and scalable resources capable of being expanded as required
- Not expensive – the function needs to provide excellent value for money
- Well trained and technically proficient
- 24 /7 (or near) Support and cover for clients and operational processes
- Technicians capable of meeting and discussing requirements with clients (and then building the necessary systems)
- Managers who can embrace and understand new legislation

Technology as a force of Change

Turning to the technology side there are significant changes coming to Data Processing Departments due to the advancement of technology. Advances in chip technology, data storage and retrieval, and communications systems allow us to model and programme more complex algorithms inside the computer. Advances in software development itself have made computer systems more stable and less bug-ridden.

Raw computing power is enabling barriers to be pushed out further. More is possible so systems can cope with greater levels of complexity. Survey collection scripts, be they Web, Cati or Capi, can be more complex. Real time x-tabulation over the web is now a possibility. Charts and presentations can be created automatically. This complexity comes with a price – we need to make sure we have the appropriate skills and tools to build and maintain these ever-growing complex systems. We also need less skilled people to operate these systems – leading to a polarisation of skills, and even some "De-skilling".

Complexity of processing functions will lead to "Black Boxes" being built and this will lead to the polarisation of skills into highly technical consultants and less technical clerical and administration staff.

The Internet continues to have an ever-increasing impact on our industry. The volume of papers relating to the internet at this conference indicate this is an area we cannot ignore when thinking about where we will be in the future.

At BMRB the Internet is having an impact on us in four ways:

- Web collection is growing as a methodology, especially when combined with other modes.
- Optimisation of certain operational functions, for example in communicating with interviewers in the field.
- Publishing and disseminating results to our clients.
- Client based "Portal" which will allow clients to monitor the progress of jobs, provide access to a Job library, download relevant files and carry out further analysis on the data in real time.

I am sure the impact of the Internet on the DP function will continue over the coming years. There will be time when all electronic data collection will be through a standard browser regardless of methodology (CAPI, CATI or Web).

The final big agent of change we are facing is that DP and IT are at last merging**. DP is finally going mainstream**. Some describe this as Open standards emerging; others describe this as Microsoft taking over the world. Regardless of which view you take it is clear that the punch card and its associated technology are on its way out. SPSSmr have Dimensions, other MR software suppliers have their own way to modernise their software offerings. I am sure we are starting to see the decline of MR

proprietary software and the punch card related formats and the increasing use of relational databases and structured programming languages. The implications for training and staff development are clear.

Summary of the Challenges

We are facing pressures from all sides including the following demands:

- Fast and timely delivery for clients
- Lower costs
- Integration into client systems
- 24/7 support
- Use the tools and formats the client wants us to use
- Scalable
- Flexible
- Project management skills
- High levels of technical competence in a variety of technical fields
- Communicate internally and externally, at all levels
- Knowledge of IT and Software development

A question: is this realistic? Can we possibly deliver what our customers demand and cope with the changing technology? My view is that if we manage the clients' expectations, then the technology can help us deliver robust and reliable solutions to client problems.

I will now discuss our plans at BMRB for dealing with this changing landscape.

3. Making the Change

So where do we start in this story of increasing business pressure and advancing technology?

Our main resource is people, that is where we must start, and you know it is not going to be easy! Staff turnover in DP has generally been a major problem over the last decade, we must avoid being in that position again.

I have defined 10 action points that address the key challenges we face. At BMRB some of these action points are well advanced others need more work.

Action point 1: Managers who manage.

It is important that we have good managers who concentrate on the task of culture change and manage through the tough issues that need to be faced. There are many issues that need monitoring and chasing. Dedicated management is essential if we are to succeed.

For some organisations this culture change may require some special attention from trainers and change management consultants.

Action point 2: Technicians who can effectively communicate with clients as well as other technicians

We need to create technical consultants who are comfortable running complex projects, talking to IT suppliers and our clients. We have some skills in this area and have nothing to fear and much to gain by involving technicians in the discussions with clients. We are yet to define competencies in this area.

Action point 3: Inexpensive, flexible, offshore resources to carry out the basic processing functions (India)

At the heart of this plan is a low cost, flexible, offshore solution for processing data. This resource needs to be flexible in terms of the volume it can cope with.

We have an offshore team of 8 – 10 people that complements the UK DP team. The India processing option works because the team there is talented (post grads), and they have a great work ethic and attitude. By the end of this year about half of the BMRB adhoc output will be processed in India.

Offshore processing is not a magic wand and requires hard work to make it succeed including:

- Relationship management
- Secondments both ways
- Training programs
- Communication Infrastructure
- New operating procedures
- Quality Control and management

Action point 4: Project Management skills to manage these different processes

DP staff need to be able to cope with the complexities of modern day DP life and be able to manage processes and projects other than the traditional spec and script writing tasks. These skills are not inherently a part of the DP skill set – so we will need to carry out some coaching. See Appendix 3 for Project Management competencies (Managing DP Work in India).

Action point 5: Technical and Managerial Career paths.

Technicians should be recognised for their own skills and should not be given management positions which need a different set of skills. We recognise both of these routes and during our staff development meetings we would review the options for our staff.

Action point 6: Continuous Professional Development (CPD)

This is a process for monitoring progress and aligning training to each individuals development needs. It is easy to pay lip service to this approach but hard to do it well –staff know the difference, so make the effort. True managers will see the value of such processes. See Appendix 2.

Action point 7. Secondments

Cross fertilisation of ideas is essential in today's complex world. Staff that know how other departments, clients, and suppliers work are invaluable. We look for opportunities to second staff and broaden their horizons. There is no need for staff to move to another company to gain new experiences. We have also had several secondments to India to work with our partners out there.

Action point 8. Peer Observation, Mentoring programs (BMRB's EPDP programme)

This is a BMRB DP created process that combines Quality Assurance, Peer Observation, Coaching and Process Optimisation. We allocate a consultant to spend time with each of our staff in turn. During this time the consultant watches and listens to the DP exec work for an hour or so every day for a week. At the end of the week there is a review meeting when non-optimal processes and non-conformances are discussed and reviewed. This leads to efficient processes and true Quality improvement.

Action Point 9: Gap Analysis Of IT Skills

Mainstream IT and software development skills cannot be turned on and off at will. So we need to start building some expertise in tools such as SQL, HTML, ASP, VB, C++ and .Net. Software developers who are not familiar with Market Research data have a huge learning curve to understand MR needs. Any DP staff who have the aptitude to learn the latest IT development tools need the encouragement to do so

We will carry out a "Gap Analysis" of what IT and software skills are present against what will be needed. This can then lead to a plan for training, recruitment and partnership building with other suppliers. For BMRB this means looking outside the DP department and making other parts of the organisation aware of the coming changes

Action Point 10: Refocus IT providers and align software and DP functions

We need excellent relationships with Software suppliers (such as SPSSmr) and other IT providers so that IT and the old style DP teams can work together to build the systems and processes we will need. Also this may mean ending some existing supplier agreements as we move to a new technology paradigm.

For BMRB we need to re-align the different technical functions and create an integrated team where IT, Software Developers and DP can work together. This reflects what is happening in the industry. We also need to review our helpdesk and support policies bearing in mind the new systems that are being developed such as the client data portal.

4. Conclusion

So over the next few years DP will undergo a significant transition. Script and spec writing skills using tools such as Quancept, Quantum, QSL, and Merlin were of prime importance, these will no longer be the only skills required. Eventually (though not for sometime yet), these tools will become obsolete.

Overseas resources will cover the simple DP tasks and local teams will need to display excellence in other skills such as project management, software development, and client communication.

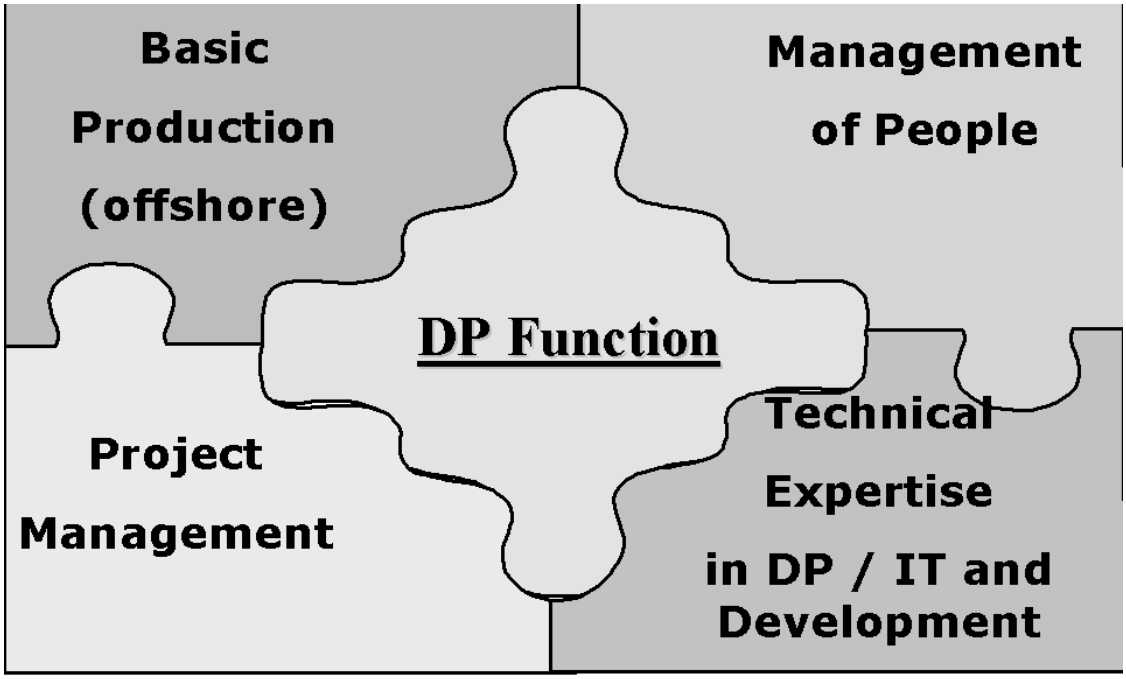
Organisational structure needs attention so technical areas are not in silos – but share knowledge and experience.

Unfortunately not all the current staff will have these new skills, or the desire and aptitude to gain them. However as these changes will take several years to complete we will have the scope to make the adjustment with relatively little pain.

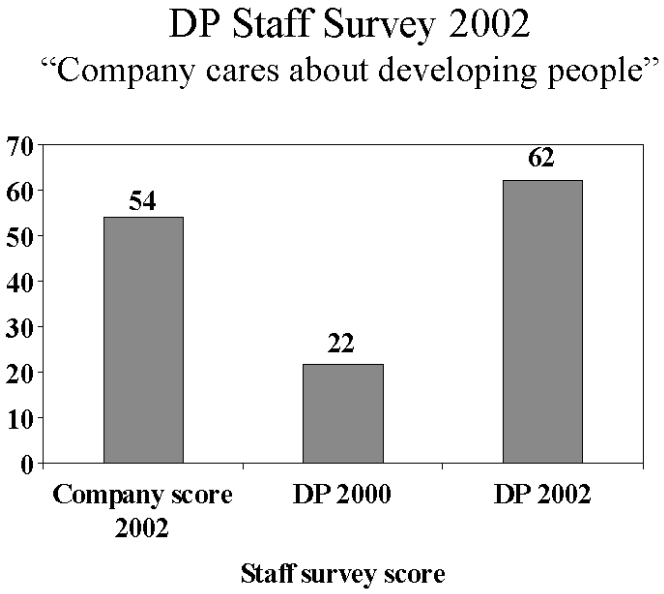
With a strong management team (which I am lucky enough to have!) we will be able to make this change. So to finish I would just like to list the essential qualities that the managers of DP will need to have in order to succeed in this new world. Once the objectives and the action plans have been developed then we will need our managers to display the following characteristics:

- Business and Financial fundamentals (in a PLC environment)
- Staff management skills (team leading, staff motivation and development, communication)
- Communication to all levels of the organisation
- Monitoring / chasing progress on key projects
- Able to sell the Culture change
- Strong technical knowledge to enable sound decisions to be made

Appendix 1. Core DP requirements



Appendix 2. DP Staff Satisfaction chart



Appendix 3 Component Definition – Data Processing

PROJECT MANAGEMENT – Managing DP Work in India	
Level 1	**Level 2**
Project specific activities	**Identifying a project as a candidate for sending to IMRB**
• Understands fully all project requirements as per Costing/Spec • Liases immediately with India, confirming deliverables and timings • Keeping schedules updated and including dates for receipt of deliverables from India • Completes Output Task Spec review • Ensure that all paper work (both electronic and/or hard copy) are maintained and that there are relevant copies in the project folder. • Communicates information effectively between research and India • Checks info from research before sending to India • Checks deliverables from India against requirements • Monitors budgets • Completes Project review/ satisfaction scores	• Costing a job • Agreeing with research potential of India doing the work • Negotiating deadlines and costs with India • Raising a Purchase Order • Maintaining effective communication with all members of project team
Negative indicators: • Spending too long solving issues rather than explaining to India how to solve • Does not fully understand project requirements • No Output Task Spec review completed • Fixes small problems rather than sending back to India	• Continually blames India for shortfalls • schedule has no dates for deliverables from India (should be negotiated beforehand) • cost Over-run

About the Author

Garj Basi is the Data Services Director at BMRB International. Prior to entering the Market Research world Garj spent four years as a programmer based in the Midlands working on large mainframe systems using languages such as COBOL, FORTRAN, and PL1.

Garj joined AC Nielsen (Oxford) in 1986 as an analyst programmer and worked on the system to process EPOS data from the major retailers. Over the next few years Garj went on to run a variety of development projects and teams. These responsibilities involved extensive travel round the world.

Garj now runs the Data Processing, Coding and Scanning functions within DP as well as being responsible for many technology based projects within BMRB. Garj has been at BMRB for nearly a decade during which time many major changes and initiatives have been implemented. These include converting the entire processing system from VAX based PTT software to the Unix based SPSSmr suite; setting up the overflow DP team in India, and recruiting around 30 graduates into the DP world.

Garj is a family man, with 3 kids and has little spare time, but manages to spend as much time as he can at his holiday home in Florida.

Garj believes the ASC is an important organisation, he said, “The world of Survey Processing will undergo some major changes over the next few years. We need to come together to share knowledge and learn from each others experiences and help navigate the wider Market Research Industry through these tough times. The ASC is an excellent vehicle for doing this.”

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Accessibility and Acceptance of a Virtual Respondent-Based Interviewer Training Application[1]

Polly P. Armsby, Michael W. Link, Robert Hubal, Curry I. Guinn, Laura Flicker, Rachel A. Caspar

Abstract

Training applications which use responsive virtual-human technology (RVHT) – training tools based on sophisticated voice recognition and behavior modeling technologies – have great potential for improving training of interaction skills essential to effective interviewing, such as refusal avoidance, probing, and addressing questions related to informed consent. However, our understanding of how to model the behavior of responsive virtual humans and how people interact with them is limited. The overall effectiveness of this technology as a training tool depends upon its ability to provide appropriate learning experiences, its ability to engage the student, and its acceptability to disparate users. This research assesses the accessibility and acceptance of a training application based on RVHT as a tool for teaching refusal avoidance skills to telephone interviewers. The assessment focuses on users' ability to understand the basic features of the application, whether diverse users are able to use the application equally, how users react to problems, whether the virtual humans are realistic enough for the users, and ultimately, whether users accept the virtual environment as a valid proxy for the real work environment.

Keywords

Interviewer training, computer-based training, responsive virtual-human technology-based training

1. Introduction

Survey research is in an era of great challenge. Response rates across all modes of data collection have been in decline, threatening the validity and utility of the information collected in surveys. As it becomes more difficult to convince sample members to participate in surveys, it is essential that the interviewers who are on the front lines of collecting these data are given the tools they need to be successful in their jobs. While interviewer training encompasses a variety of topics (CATI/CAPI skills, gaining cooperation, respondents' rights, questionnaire administration, etc.), training tools built using responsive virtual human technology (RVHT) hold the promise of offering interviewers a simulated, realistic environment for developing and practicing basic interviewing skills – such as gaining respondent cooperation, probing, administering informed consent – and honing those skills

[1] This work was supported by a research grant from the National Science Foundation (Grant No. EIA-0121211).

over time. RVHT reduces the amount of learning that must occur on the job, by allowing repetitive practice in a virtual environment.

RVHT is admittedly in its developmental infancy and requires additional improvements before it can be deployed as a fully mature technology in a production environment. The research presented here is one small part of a larger research program shepherding the growth and development of these technologies. RVHT involves the use of natural language processing and an emotive behavioral engine to produce natural, interactive dialogues with intelligent, emotive virtual-reality (VR) agents. RVHT has great potential for use in training interaction skills, such as those required for effective survey interviewing. However, our understanding of how people interact with responsive virtual humans (a.k.a. intelligent agents) is quite limited.

Better understanding requires employing RVHT in training applications and conducting systematic use, usability, perception, and training-effectiveness assessments. Important questions yet to be answered include: determining whether intelligent agents make learning more accessible; determining whether students are willing to accept intelligent agents as interactive partners in learning; determining what skills can be acquired, practiced, and validated using RVHT; determining what is involved in providing a convincing simulation of human interaction, realistic enough for the student to suspend disbelief and acquire skills that will transfer to a "live" environment.

Users' interactions with RVHT applications are little studied and poorly understood. The research presented here (and the larger research program from which it is drawn) provides an initial assessment of some of the issues associated with user interface design, user acceptance of computer-based training, and perceptions of the realism and effectiveness of the training tool. The assessment of these issues was conducted in two phases. The first phase was conducted within a controlled environment using a dozen subject matter experts as test subjects and collected repeated assessments of the perceived performance and realism of the virtual training environment across several different hardware platforms. The second phase of the assessment was conducted within a live production environment, using approximately 50 telephone interviewers of varying backgrounds to test the application. A longer, structured questionnaire was then used to capture their evaluations and perceptions of the training tool. Both assessments involved the use of an RVHT-based training tool for refusal avoidance at the outset of a telephone interview. The findings seem to indicate that RVHT has great promise as a tool for training survey interviewers; however, the underlying technologies need further development before such applications are robust enough to be fully production-ready.

2. Mechanics of the RVHT Survey Interviewer Training Application

Successful interviewers must employ a variety of skill sets including standardized interviewing practices, proficiency using a computer and relevant software, a thorough knowledge of the current survey instrument, and interpersonal, interaction and active listening skills. Research has shown that flexibility is critical for developing effective interaction skills (Groves & Couper, 1998) and for performing well under time constrained, information-poor, and other difficult conditions (Klein, 1998). In order to acquire flexible and effective approaches to gaining respondent cooperation, new and experienced interviewers require a learning environment that realistically simulates the environment they face in an interviewing situation. The consistency that is gained by repetitive practice in virtual and constructive learning environments leads directly to effective decisions in the production environment (Ross, Pierce, Haltermann, & Ross, 1998). Practice also leads to increased confidence before the first real on-the-job experience, minimizing the amount of on-the-job learning

that is necessary. In the survey world, on-the-job-learning can translate into numerous unsuccessful interview attempts at the start of a study by a new interviewer, leading to lower response rates, lower quality data, delayed schedules, and increased costs.

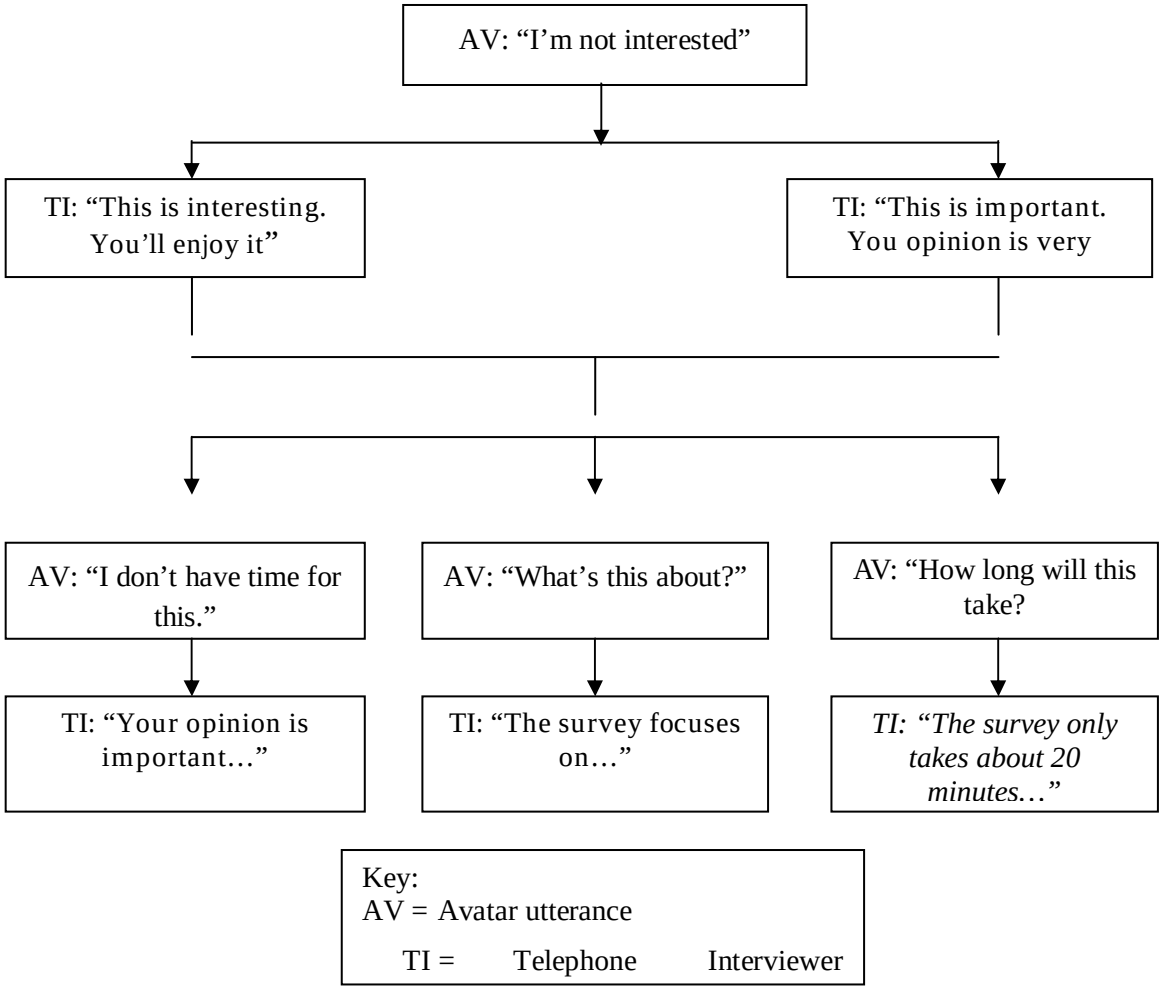
This is exactly the type of scenario in which RVHT can be most effective. The outset of any interview is generally very fluid, despite the fact that interviewers are nearly always provided with an introductory script or set of bullet points for making the introduction. Sample members often interrupt interviewers with a barrage of questions or remarks, such as "I'm not interested," and "I don't have time." Non-response research suggests that the best approach to obtaining participation is for the interviewer to immediately reply with an appropriate, informative, tailored response (Camburn, Gunther-Mohr, & Lessler, 1999; Groves & Couper, 1998; Groves, 2002). Generally, such skills are taught through a combination of lecture, paired-mock practice with other interviewers, and by using multimedia to listen to real or mock audiotapes of exchanges between interviewers and sample members. RVHT allows us to take skill building to the next level, by providing a realistic, simulated environment in which an interviewer can practice and hone his or her skills.

The application tested here involves the use of an RVHT-based application to simulate the environment a telephone interviewer faces during the first thirty to sixty seconds of a telephone survey interaction. The training tool allows interviewers to practice their skills in gaining cooperation in a self-paced, realistic environment. The software is designed such that interviewers begin with an introduction and are then required to respond to a series of objections and questions raised by the "virtual respondent." The interviewer's responses are captured electronically and processed by a natural language speech processor. Based on the content of the interviewer's speech, the software launches another objection/question or ends the conversation by either granting the interview or hanging-up the telephone (see Figure 1).

The application uses spoken natural language interaction (Guinn & Montoya, 1998), not text-based interaction (except for data collection during development from subject-matter experts). The speech recognizer uses a basic dictionary of common words as well as a specific dictionary for each turn of a conversation. The specific dictionary consists of up to 200 words based on behavioral observations of real world events. These specific dictionaries are dynamic, therefore, changing with each turn of the conversation.

The application tested here is not designed to present training content to the user. The basic skills needed for gaining cooperation are presented and initially acquired during an instructor-led classroom session. However, one of the primary potential benefits of the application is the enhanced ability to practice these critical skills using repetitive, structured, and standardized scenarios and to conduct practice sessions outside of the traditional classroom setting. While RVHT applications can be used to direct content to the trainee, this research focuses primarily on the practice component of the training curriculum. Finally, the current system is not yet equipped to provide intelligent tutoring. That is, coaching and feedback (aside from an indication of success and failure) are provided by direct observation from supervisors and trainers during or after use of the application, rather than from the application itself.

Figure 1. Example of Dialogue Flow



3. Assessment 1: Controlled Test Environment

Overview

The primary purpose of our first assessment was to capture users' evaluations of the overall performance and perceptions of realism of the application across several different computer platforms. Assessing gains in the persuasion skills of the test subjects was not within the scope of this (or the second) assessment; rather the focus is on perceived performance and perceptions of realism of the application itself. The experiment consisted of testing four hardware platforms with twelve volunteer subjects who were blinded to the machines being tested. Each test consisted of three separate conversations with the Avatar on the selected hardware platform. After completing those three conversations, the subject was asked to rate the experience with respect to realism of the simulation using a three-question assessment form.

Analysis Measures

Analysis variables were derived from two sources: (1) coded responses from the transcripts of interactions between the Avatar and the subject and (2) evaluations made by the subjects themselves.

Transcript-derived Measures

The taped conversations were first transcribed (with the transcriptions being verified by the test administrator). Then each conversation was coded to indicate unique conversation exchanges and the semantic meaning or focus of each exchange. In all, there were a total of 910 unique exchanges that were coded from the 264 conversations (which represent 88 different trials across the four machines). From the coded transcripts, three measures were developed to measure the behavior of the RVHT application:

- **Conversation Exchange:** measures the number of Avatar-subject conversational interactions. An "exchange" is defined as the pairing of an Avatar "objection" and a subject "response."
- **Conversation Semantic:** measures the content or meaning of the exchange between the Avatar and the subject. Initially all exchanges were coded into one of 35 possible "semantic" categories. These 35 categories were then collapsed into six general conversation semantics: Introduction, Survey Content, Time Concerns, Selection Criteria, Survey Attributes, and Setting
- **Conversation Complexity:** measures the number of unique semantics observed during the course of a conversation. A conversation with a larger number of unique semantics is considered to be a more "complex" conversation than one with fewer unique semantics.

Subject-derived Measures

Three additional measures were developed from observations made by the subjects themselves. For each machine trial, subjects completed three separate conversations with the Avatar. After each set of conversations the subjects were asked to rate the realism of the trial in terms of responsiveness, overall conversation, and the objections raised. Each of these dimensions was rated on a seven-point scale, where 1 = not at all realistic and 7 = extremely realistic.

- **Realism of Response Times:** Did the application respond quickly enough to mirror the way in which sample members actually respond over the telephone?
- **Realism of the Overall Conversations:** Did the dialogue that took place during the three conversations generally reflect the types of dialogues (in terms of flow and content, pace and tone) that take place with sample members at the outset of a telephone interview?
- **Realism of the Objections Raised:** Were the objections raised by the Avatar realistic and reflective of those encountered in exchanges with reluctant sample members during actual interviews?

Findings

The analysis was conducted in two parts: (1) subject evaluations of the application's performance across four platforms and (2) evaluation of the relationship between subjects' evaluations of realism and the behavior of the application in terms of exchanges, semantics, and complexity of the conversations.

Subject Rating of Application Realism by Platform

First, we examined how the subjects themselves rated their practice experiences across the four platforms. The unit of analysis in this section is the "trial" level (i.e., a trial equals three conversations conducted on a single machine). In all there were 88 trials conducted across the four machines. In terms of evaluating the realism of the response times across the platforms, the differences (while not statistically significant at the traditionally excepted level of $p < .05$) are suggestive of a significant difference (given the relatively small sample size of 88 and a $p < .096$ value). In terms of response time, the Dell laptop rated the highest (5.06 average rating), followed by the IBM ThinkPad (4.87), the

Gateway PC (4.38), and the IBM PC (3.96). The laptops, therefore, ranked higher than the PCs in terms of subjects' ratings of their response time.

Subjects demonstrated little difference in their evaluations of the other two dimensions of realism – assessment of the overall conversation and the objections raised. There were no significant differences noted across each measure in terms of platform used or testing trial. Outside of possible differences in perceptions of response time, therefore, subjects found little difference in their ratings of the realism of the practice conversation generally and of the specific content of those conversations. Likewise, there appeared to be little "educating" of the subjects between trials 1 and 2.

Subject Rating of Realism Based on Application Behavior

Our second area of interest focuses on how subjects' ratings of realism may have been affected by the behavior of the application itself. We might expect that if the application is perceived to behave in a more "realistic" way that we should see differences in subjects' ratings of response time, conversation flow, and content. We found, however, surprisingly few differences in ratings on these dimensions across the different measures of Avatar behavior (conversation exchanges, semantics, and complexity). Again, the results presented here are at the "trial" level.

There were no significant differences seen across these three dimensions based on the average number of exchanges per conversation within a trial. Trials with an average of 1.0 to 2.9 exchanges per conversation were not rated significantly higher or lower in terms of response time than those with 4.0 to 5.0 exchanges. The same is true when we look at ratings of the overall conversation and the objections raised during the exchanges.

Likewise, there was little variation across the three realism dimensions when we consider the six general conversation semantics. The only statistically significant difference was noted in terms of evaluation of the realism of the objections raised. When "setting a callback" was a topic of a trial, that trial tended to be rated higher in terms of the realism of the objections made, than did trials where setting a callback was not a focus.

Finally, and somewhat surprisingly, the complexity of the conversations across a trial was not related significantly to ratings of response time, nor of the realism of the overall conversation and objections raised. One might have expected that interactions that are more complex would have led to higher ratings on either or both the overall conversation or objections raised dimensions. This, however, was not the case.

Summary of Controlled Environment Assessment

In terms of user perceptions, there were few notable differences discerned. Subjects did not vary significantly in their evaluations of the realism of the response time, overall conversation, or objections raised across different platforms and trials for the experiment, nor across differences in the types of exchanges they encountered (shorter/longer, more/less semantically complex). In part, this may be due to the low number of observations resulting from this analysis being conducted at the trial-level (the trial level was used for analysis since that is the level at which the perception evaluations of realism were made). Further analyses will be conducted using more sophisticated statistical modeling (nested data analyses) at the conversation and exchange levels to determine if significant differences in perceptions are revealed at those levels.

4. Assessment 2: Operational Test within Live Production Environment

Overview

A primary goal of the overall research program of which this study is a part is to determine if RVHT can be an effective technology for interaction training across a broad spectrum of ethnic and socio-economic backgrounds, jobs, and job levels. The effectiveness of this technology depends upon its ability to provide appropriate learning experiences, its ability to engage the trainee, and its acceptability to disparate users.

In the second phase of our assessment, we examine this aspect by collecting data from a group of approximately 50 telephone interviewers of varying ages, races, experience and education levels, who used the refusal avoidance training module within a production environment. To evaluate the accessibility of the application we focused on the following:

- Do users understand the basic features of the application?
- Are users able to complete each task and exit the application?
- Are different users (e.g., based on ethnicity, job level, and education level) equally able to use the application?

Analysis of these questions will provide clues as to how smoothly the application runs, or when and why difficulties arise in its use.

The question of whether and why participants "accept" or "reject" the virtual training environment is also central to this research. To evaluate acceptance of the application by the trainees, we debriefed participants using a structured questionnaire and moderator-facilitated focus groups to gauge reactions and engagement in the application. In particular we are interested in the following:

- Are the virtual humans realistic enough for the users? Why or why not?
- How fast and accurate is the speech recognition?
- Could trainees detect changes in the emotive states of the virtual human using only audio cues?
- Would they use the application again and/or recommend its use by others?

As part of this second phase of the evaluation process, data were collected using a questionnaire filled out by the interviewers and notes made by instructors and researchers who observed the training sessions. The questionnaire asked questions related to users' perceptions of the realism of the interactions with the "virtual human," ease of use of the software, the perceived effectiveness of the training sessions, and some basic background characteristics of the users. In all, a diverse group of 48 interviewers filled-out the questionnaires (96% of the software users).

Findings

The questions posed to the interviewers were designed to assess their perceptions and experiences in using the RVHT training tool in four basic areas: ease of use of the software, realism of the training environment, impact on skill development, and desire to recommend or use the software again. Although this is the first detailed look at how users interact emotive intelligent agents for soft-skills development, we can formulate some hypotheses regarding how different types of users might respond based on how users generally differ in their use and acceptance of other computer-based tools. For example, we might expect to find that trainees who are younger, have more education, and are more comfortable using computers in general to have fewer difficulties in using the system. Likewise, we might expect that more experienced interviewers might not find the training tool as useful as inexperienced interviewers because the more experienced interviewers will have already developed

and honed their refusal avoidance skills (a supposition that mirrors the finding of Groves, 2002). To examine possible differences in accessibility and acceptance of the program, we cross-tabulated all of the closed-ended questions in the questionnaire with demographic characteristics including sex, education level, age, race, and work experience. Significant differences are noted below.[1]

Ease of Use of the Application

Users of the RVHT software seemed to find it very accessible to use, with 84% indicating the software was either extremely easy or very easy to use (52% extremely, 31% very, 13% somewhat, 4% not too, 0% not at all). Nearly everyone found the written instructions (96%) and the verbal instructions (98%) that accompanied the training to be clear and accurate. Only eight (17%) of the 48 trainees indicated that they required additional assistance to use the training software (after the initial training received by all trainees).

Realism of the Training Environment

The promise of RVHT-based training tools is that they can simulate a "real" environment, thereby allowing trainees repetitive practice in conditions that are as close as possible to what they will encounter on the job. For this particular application, the "virtual respondent" needed to mirror the behaviors and emotions of real respondents encountered when doing live interviewing. This means delivering an array of objections to the trainees in different tones of speech and emotional levels in a fast-paced manner. Interviewers were asked a series of questions to try to assess how well they accepted the virtual environment as a substitute for real work conditions. In other words, do they "buy-into" the virtual environment?

The answer is somewhat mixed. In general, trainees did not find the virtual environment to be realistic and they cited two primary reasons: the slowness of the response of the "virtual respondent" and the limited number of different objections/questions offered by the "virtual respondent." They did, however, find the responses that were offered to be realistic and stated that they could detect and respond to changes in tone and emotional cues offered by the "virtual respondents." A majority of the trainees also indicated that they felt the sessions helped them to improve their skills needed at the outset of an interview either somewhat or a lot.

When asked, In general, how realistic did you find the overall conversation with the 'virtual respondent,' 17% said they thought it was extremely or very realistic, 44% said it was somewhat realistic, 17% not too realistic and 23% not at all realistic. Slowness of the "virtual respondents" in replying (due to the lag caused by the speech recognizer as it interpreted the interviewer's responses and determined the next script to launch) was the primary problem cited by interviewers. Perhaps not surprisingly, trainees who describe themselves as "fast-touch typists"were more likely than those who indicated they were "slow-touch typists" to say the response time was too slow (82% fast-touch vs. 67% slow-touch; $p < .08$ chi-sq.).

The trainees were, however, more positive when evaluating the realism of the objections and questions offered by the "virtual respondent." A plurality (48%) indicated that the content of what was said was either extremely or very realistic, with 40% saying it was somewhat realistic, 8% not too realistic, and

[1] Because of the small number of observations (N=48) we also created dichotomous variables for both the dependent variables (collapsing scales where possible) and independent variables (collapsing or combining variables with 3 or more values). These variables were also examined to determine if significant differences among subgroups could be identified.

4% not at all realistic. They also felt it was relatively easy to determine the emotional state of the virtual respondent based on the tone of voice they heard (23% extremely easy, 44% very easy, 29% somewhat easy, and 4% not too easy; no one indicated that they could not determine the avatar's emotional state from the tone of the "virtual human's" voice). Likewise, the content of the speech used by the avatar was also a good cue to trainees as to the "virtual human's" emotional state: 8% extremely easy to tell, 54% very easy, 27% somewhat easy, 10% not too easy, 0% not at all easy.

Nearly 60% indicated that they behaved differently in the practice scenario based on the tone of the virtual respondent's voice. Interestingly, a higher percentage of women than men reported reacting differently to the changing tone of the avatar's voice (women 67% v. men 33%, $p < .04$ chi-sq.). Similarly, 54% said they treated the situation differently based on the actual words used by the avatar in expressing a concern or voicing an objection. There were, however, no differences between men and women on this question. It seems, therefore, that the both the content of the objections raised by the virtual respondent and the emotional behavior of the "virtual human" were generally accepted by the trainees and caused them to react differently within the various training scenarios.

Impact on Skill Development

The purpose for allowing trainees to operate within a virtual environment is to allow them to develop and hone essential skills before entering the "real" environment, thereby reducing the amount of "on the job" skill development required. New interviewers can do considerable damage at the outset of a telephone study, generating a large number of refusals as they gain comfort and confidence on the telephone. If practice within a virtual environment at the beginning of a project can reduce the numbers of initial refusals even modestly, then the training program will have value. While longer-term assessments of the effectiveness of the RVHT software will need to include examination of more objective measures of improved performance, this preliminary assessment focused on the user's assessment of the impact of the training on their own skill development.

Trainees were asked to evaluate if they thought the RVHT software increased their abilities in six different areas. Nearly three-quarters of the trainees felt that the practice sessions increased a lot or somewhat their ability to respond to questions and concerns by sample members. Approximately 56% felt it helped them a lot or somewhat in better gaining respondent cooperation at the outset of an interview. Likewise, over half felt it helped in their ability to adapt to differences in respondents' tone or voice or perceived moods and to adapt to differences in the speed and pace of different sample members' speech. About half of the trainees also thought that the sessions helped them a lot or somewhat in avoiding refusals at the outset of an interview.

Would They Use the RVHT Training Tool Again?

An effective training tool is also one that trainees should enjoy using, would use again, and recommend to others). Approximately two-thirds (65%) of the users said that they found using the RVHT software to be fun and enjoyable. Interestingly men were significantly more likely than women to say that they found the sessions to be enjoyable (92% men vs. 56% women, $p < .05$ chi-sq.). Nearly three-quarters (73%) said they would like to use the software again. In addition, 83% said they would recommend the program as a training tool for other interviewers. In open-ended responses, a number of interviewers indicated that it would be a very good practice vehicle for new or less experienced interviewers.

Summary of Live Environment Assessment

This initial assessment of an RVHT-based training tool for telephone interviewers provides some valuable insights into how trainees access and accept virtual environments as practice labs and "virtual humans" as training partners. There were aspects of the training program that interviewers clearly liked, such as the ability to do repeated practice of frequently asked questions, being able to distinguish different emotional states from the tone of voice and speech content of the virtual respondent, and the opportunity to learn to think on their feet in a simulated environment before being placed into a live interviewing situation.

There were also aspects that the interviewers did not like, such as the slowness of the response of the virtual respondent and the perceived lack of variety in the scenarios that were presented. This provides constructive feedback for the engineering and improvement of the software. While adding additional scenarios is a relatively easy process, involving research into the "normal" flow of such scenarios , the responsiveness issue is a more fundamental matter, reflecting the current state-of-the-art in speech recognition. For virtual training partners to be more readily accepted, the underlying speech recognition technology needs to be improved, providing, faster processing of the input from interviewers and launching of responses by the virtual respondent.

5. Conclusions

A considerable amount of basic research is still required to make RVHT applications robust, viable training tools within production environments. RVHT can hold one of the keys, however, for improved training of interviewers – both telephone and field-based staff. The research provided here offers additional information allowing developers and application designers a greater understanding of how RVHT applications respond under repeated test conditions and will hopefully help speed the development of these much needed training tools.

We do not anticipate RVHT-based training will replace instructor-led training, but we expect that combinations of RVHT-based training and instructor-led training will continue to offer advantages for presenting training exercises that are more uniform and realistic than those that can be reproduced in the classroom alone. Additionally, RVHT-based training can provide easily implemented, focused sustainment (i.e., refresher) training.

Future research will continue to examine if the presence of the RVHT applications allows trainers to increase the amount of time spent on skill acquisition, by reinvesting time spent on individual practice into classroom sessions focused on existing or additional content. Now that initial assessments are complete, efforts can be made to improve the realism of practice sessions by compiling a more robust corpus of Avatar objections, thus allowing the user to engage in a richer conversation with the application. In an effort to address technical issues that may detract from the realism or the reliability of the training application, further testing and analyses will be conducted in an effort to determine the source of variability in application behavior across hardware platforms. Furthermore, we hope to make improvements to the engineering of the application itself as the underlying speech recognition technology improves. Last, in future experiments involving more rigorous testing with a larger group of users and including controlled experiments comparing users' and non-users' performance, we hope to discern whether an RVHT application contributes to learning beyond what is offered through classroom, multimedia, and computer-assisted instruction.

We feel it is important to continue to investigate more robust and effective RVHT models and more efficient means of creating the models, to better understand user preferences and acceptance of RVHT, and to determine how best to use RVHT in combination with other training methods to provide cost-effective training on critical interaction skills.

References

Bates, J. (1994). The Role of Emotion in Believable Agents. Communications of the ACM, Special Issue on Agents, July, 1994.

Becheiraz, P., & Thalmann, D. (1998). A Behavioral Animation System for Autonomous Actors Personified by Emotions, Proceedings of the First Workshop on Embodied Conversational Characters (WECC '98), Lake Tahoe, California.

Camburn, D.P., Gunther-Mohr, C., & Lessler, J.T., (1999). Developing New Models of Interviewer Training. International Conference on Survey Nonresponse, Portland, OR, October 28-31, 1999.

Cassell, J., & Vilhjálmsson, H.H. (1999). Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous. Autonomous Agents and Multi-Agent Systems: 2, 45-64.

Cole, R., et. al. (1999). New tools for interactive speech and language training: Using animated conversational agents in the classrooms of profoundly deaf children. In Proceedings of ESCA-MATISSE ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education, London, UK.

Elliott, C. (1993). Using the Affective Reasoner to Support Social Simulations. In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, pp. 194-200, Chambery, France, August 1993. Morgan Kaufmann.

Graesser, A., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the Tutoring Research Group (2000). AutoTutor: A simulation of a human tutor. Journal of Cognitive Systems Research, 1,35-51.

Gratch, J. (2000). Modeling the Interplay Between Emotion and Decision-Making. Proceedings of the Ninth Conference on Computer Generated Forces & Behavioral Representation, May 16-18, 2000, Orlando, FL.

Groves, R. (2002). Principles and Practices in Non-response Reduction. Presentation at the 2002 Respondent Cooperation Workshop sponsored by the Council for Marketing and Opinion Research, New York, NY.

Groves, R., & Couper, M. (1998). Nonresponse in Household Interview Surveys. New York: John Wiley & Sons, Inc.

Guinn, C.I., & Montoya, R.J. (1998). Natural Language Processing in Virtual Reality, Modern Simulation and Training, pp. 44-55.

Helms, R.F., Hubal, R.C., & Triplett, S.E. (1997). Evaluation of the Conduct of Individual Maintenance Training in Live, Virtual, and Constructive (LVC) Training Environments and their Effectiveness in a Single Program of Instruction. Final Report, September 30, 1997. Submitted to Battelle RTP Office, Subcontract # TCN 97031, Delivery Order #0027, Dated April 16, 1997.

Hubal, R.C., Kizakevich, P.N., Guinn, C.I., Merino, K.D., & West, S.L. (2000). The Virtual Standardized Patient-Simulated Patient-Practitioner Dialogue for Patient Interview Training. In J.D. Westwood, H.M. Hoffman, G.T. Mogel, R.A. Robb, & D. Stredney (Eds.), Envisioning Healing: Interactive Technology and the Patient-Practitioner Dialogue, 133-138. IOS Press: Amsterdam.

Klein, G. (1998). Sources of Power. MIT Press.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., & Bhogal, R. (1997). The Persona Effect: Affective Impact of Animated Pedagogical Agents. In S. Pemberton (Ed.), Human Factors in Computing Systems: CHI'97 Conference Proceedings. 359-366. New York: ACM Press.

Loyall, A.B., & Bates, J. (1997). Personality Rich Believable Agents That Use Language. Proceedings of the First International Conference on Autonomous Agents, February 1997, Marina del Rey, CA.

Lundeberg, M., & Beskow, J. (1999). Developing a 3D-Agent for the August Dialogue System. Proceedings from AVSP '99, Santa Cruz, CA.

Maggart, L.E., & Hubal, R.C. (1998). A Situation Awareness Model. In S.E. Graham & M.D. Matthews (Eds.), Infantry Situation Awareness: Papers from the 1998 Infantry Situation Awareness Workshop. U.S. Army Research Institute: Alexandria, VA.

Massaro, D.W., Cohen, M.M., Beskow, J., Daniel, S., & Cole, R.A. (1998). Developing and Evaluating Conversational Agents. In Proceedings of Workshop on Embodied Conversation Characters (WECC), Lake Tahoe.

Psotka, J. (1995). Immersive Training Systems: Virtual Reality and Education and Training. Instructional Science 23, 405-31.

Rousseau, D., & Hayes-Roth, B. (1997). Improvisational Synthetic Actors with Flexible Personalities. KSL Report #97-10, Stanford University.

Shlechter, T.M., Bessemer, D.W., & Kolosh, K.P. (1992). Computer-Based Simulation Systems and Role-playing: An Effective Combination for Fostering Conditional Knowledge. Journal of Computer Based Instruction 19, 110-114.

Shneiderman, B. (1992). Designing the User Interface: Strategies for Effective Human-Computer Interaction. Reading, MA: Addison-Wesley.

Weiss, E. (1993). Making Computers People-Literate. San Francisco: Jossey-Bass.

About the Authors

Polly P. Armsby is a Survey Director at RTI International. She coordinates tasks such as design, implementation, and data collection and plans data collection schedules, supervises data collection activities, manages data quality control, trains data collection staff, and conducts methodological research. She can be contacted at RTI International, 3040 Cornwallis Rd., P.O. Box 12194, Research Triangle Park, NC 27709-2194, USA; tel. (919) 248-4514, fax (919) 316-3863; e-mail: armsby@rti.org.

Michael W. Link, Ph.D. is a Senior Survey Director and Methodologist at RTI International. He directs and manages survey-related projects and conducts methodological research in the areas of survey operations, interviewer training, and uses of new technology in data collection efforts.

Robert Hubal, Ph.D. is a Senior Research Engineer at RTI International. He is responsible for exploring technology assisted learning methodologies and identifying approaches to improving learning and training effectiveness with an emphasis on cost-effective, technology-enhanced innovations in the design, development, delivery, evaluation, and management of training and education.

Curry I. Guinn, Ph.D. is a Research Computer Engineer at RTI International. He leads the development of interactive virtual humans with emotions that affect their body and facial gestures, decision-making, and language generation.

Laura Flicker is a Survey Methodologist at RTI International. She provides expertise in survey design and implementation and has helped develop many original survey instruments on a wide range of topics including healthcare, drug abuse, welfare, education, housing, and transportation.

Rachel A. Caspar is a Senior Survey Methodologist with RTI International. She specializes in designing questionnaires and developing data collection procedures for surveys of sensitive topics, has conducted methodological studies to improve data quality, and has conducted studies of both unit and item survey nonresponse.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

First Steps along the Audit Trail

Tim Burrell

Abstract

ONS uses a software named Blaise for data collection. In the last year we have started to examine the potential of the audit trail facility within Blaise. An audit trail records all the interactions on the laptop. These interactions could be made by the interviewer or the respondent during computer assisted self-interviewing (CASI). The audit trail can provide a detailed history of the sequence and timing of navigation, field entries, changes to field entries and other events.

With the help of a package provided by Statistics Canada named Atlas, we have been able to enhance our use of the audit trail facility and widen the extent to which we have implemented it on various surveys. This new software has also led too much easier analysis of the extensive data produced.

The paper will outline the practical application of Audit Trails on two projects: 1) to monitor respondent behaviour during CASI; and 2) to observe interviewer actions on a survey of income and living conditions. Audit Trails have proved useful in both the timing and testing of questionnaires as well as trying new procedures and ways of carrying out data collection.

The paper will describe how we have been able to analyse the data produced, with reference to surveys mentioned above, and what progress we would like to make in the future. It will also consider some of the implications and concerns we have on using this tool.

Keywords

Blaise; audit trails; ATLAS

1. Introduction

In the last few years the Office for National Statistics (ONS) has started to examine the potential of the Blaise Audit Trail facility. An Audit Trail produced within Blaise records interactions made by the interviewer during computer assisted interviewing (CAI) or the respondent during computer assisted self-interviewing (CASI). The Audit Trail can provide a detailed history of the sequence and timing of navigation, field entries, changes to field entries and other events.

The paper will outline the practical application of Audit Trails on two projects:

1. to monitor respondent behaviour during CASI on the General Household Survey (GHS); and
2. to observe interviewer actions on a survey of income and living conditions (EU-SILC).

The paper will describe how we have been able to analyse the data produced, with reference to surveys mentioned above, and what progress we would like to make in the future. It will also consider some of the implications and concerns we have on using this tool.

With CAI, and Blaise in particular, the industry standard for the collection of official survey statistics, there is an increased availability of data for the researcher to investigate interviewer and respondent behaviour during an interview. With the help of a package provided by Statistics Canada named ATLAS, we have been able to enhance our use of the Blaise Audit Trail facility and widen the extent which we have implemented it on surveys. This new software has also led too much easier analysis of the extensive data produced.

2. Previous use of the Audit Trail within ONS

ONS has previously used the Blaise Audit Trail facility on a small-scale project to evaluate the utility of this monitoring method. We chose an ONS survey of children and adolescents in the care of local government authorities, known as the survey of Looked After Children (LAC) as the vehicle for this trial. The Blaise instrument for the survey was designed to collect information about the mental health of children and included a substantial CASI section relating to sensitive or illegal behaviour. The main interest of the trial was to monitor respondent behaviour in the Audio Computer Assisted Self-Interview (A-CASI) section of the interview. (Bumpstead 2001)

This study confirmed in ONS earlier findings (Bumpstead 2001, Hansen and Marvin 2001) that the Blaise Audit Trail can provide useful data about respondent use of CAI instruments which it would not otherwise be possible to capture. The processing and analysis of audit trail data is not generally straightforward. However, it was possible to formulate a strategy for dealing with audit trail data which was not unduly cumbersome or time consuming.

3. ATLAS - A tool provided by Statistics Canada

ATLAS has been used to aid in the analysis of the vast amounts of data produced within an Audit Trail. It has the ability to read and average out across cases the times of individual questions or sets of questions from the Blaise Audit Trail. This proved useful as we could separate out the time spent in particular parts of the interview, such as the interview section of the questionnaire and in the administrative section. Measurement of elements of the interview was the main point of the projects mentioned in this paper.

Before acquiring the ATLAS package from Statistics Canada, ONS had only been able to carry out limited analyses of Audit Trail data. We had looked at information case by case. Data was produced like that shown in Figure 1, which shows the details from one case. This type of data is hard to read and Audit Trail data is hard to analyse in its raw form. Figure 1 shows the interviewer's progress through 4 questions. In this particular case, the interviewer took 11seconds to pass through this set of questions.

Figures 2 and 3 show how the information is presented using the ATLAS. Figure 2 shows information from the same case as Figure 1. Figure 3 shows the average time over a number of cases (153 cases from the UK's General Household Survey, GHS). It shows that, on average, interviewers took 7 seconds to pass through these questions.

Figure 1. Audit Trail raw data

```
"19/06/2002 18:16:31","Enter Field:QSignIn.StartDat","Status:Normal","Value:"
"19/06/2002 18:16:37","Leave Field:QSignIn.StartDat","Cause:Next
Field","Status:Normal","Value:20020619"
"19/06/2002 18:16:37","Enter Field:QSignIn.DateChk","Status:Normal","Value:"
"19/06/2002 18:16:40","Leave Field:QSignIn.DateChk","Cause:Next
Field","Status:Normal","Value:1"
"19/06/2002 18:16:40","Enter Field:QSignIn.IntEdit","Status:Normal","Value:"
"19/06/2002 18:16:41","Leave Field:QSignIn.IntEdit","Cause:Next
Field","Status:Normal","Value:1"
"19/06/2002 18:16:41","Enter Field:QSignIn.WhoHere","Status:Normal","Value:"
"19/06/2002 18:16:42","Leave Field:QSignIn.WhoHere","Cause:Next
Field","Status:Normal","Value:1"
```

Figure 2. Analysing the Audit Trail using ATLAS for one case

```
==============================================================
         INITIAL          SUBSEQUENT         TOTAL
BLOCK    HITS    SECS     HITS      SECS     HITS       SECS
QSIGN     1       11       0         0        1          11
```

Figure 3. Analysing the Audit Trail using ATLAS for a number of cases

```
=====================================================================
         INITIAL          SUBSEQUENT         TOTAL      AVG
BLOCK    HITS    SECS     HITS      SECS     HITS       SECS
QSIGN     4        7       0         0        4          7
```

It can be seen from the screen shots above that ATLAS summarises data in an aggregated form which is much more convenient as well as being in a much more readable form . This has enhanced our ability to analyse data produced by the Audit Trail. The data can be shown at the case level (figure 2) or the average over a number of cases (figure 3).

The data produced from ATLAS can be easily transformed into SPSS format where data can be analysed further. Other information available in the tool was the ability to single out particularly long hits, or unusually long amounts of time spent on questions. This feature means it is possible to single out any questions where a very long time was taken for an answer to be filled out. This could be due to a break in an interview for a number of reasons. It is then possible to remove such an outlier from the analysis if it would bias the measurement of the overall time spent on that question.

Data can be analysed both at an individual level, looking at the data case by case, or by looking at averages over all cases. Additional features include the ability to look at the number of fields which were answered *don't know* and *refusal*, and changes or edits to questions.

4. A pilot to monitor respondent behaviour during CASI on the General Household Survey (GHS)

The GHS is a multi-purpose survey providing the government with annual information about the major social fields of Population, Housing, Employment, Education, Health and Income to supplement the more specialised best national sources for estimates on these topics, such as the Labour Force Survey. Because all these topics are covered in one survey, it is possible to examine the relationships between them.

The General Household Survey interview comprises two parts: the household and individual questionnaires. The household questionnaire contains questions on demographic characteristics of

household members, tenure and accommodation, consumer goods, migration and ethnicity. Questions relating to income are asked in the individual questionnaire on the GHS and the results aggregated for the household. ATLAS has been helpful in measuring the amounts of time spent in these different sections of the questionnaire.

The GHS carries topics which are asked to all members of the household. Some of these topics, such as smoking, drinking and contraception are sensitive, particularly for adolescents if there is a risk of being heard by their parents and other family members. Due to the sensitive nature of these topics they have always been carried out as self-completion sections. In previous years the self-completion sections have been entered onto paper by the respondent and then keyed in by the interviewer at a later stage. This paper-based method has the advantage, compared with CASI, that it allows all members of the household to fill out their self-completion forms at the same time rather than in turn, passing the laptop to each person.

It is possible that some information was being withheld due to this method of data collection. The respondent may see it as less confidential than if they key the data into the laptop themselves. A pilot study was undertaken to find out how long respondents would take to complete these questions by a CASI section within the Blaise questionnaire.

A pilot study was conducted out using CASI for sections of the questionnaire which previously had been completed through the paper form. The Audit Trail was enabled on the questionnaire to allow these sections to be timed. These sections included smoking, drinking, family information and contraception. Interviewers had stated that using the paper form usually took between 1 and 4 minutes for a household to complete. Self-completion also provided a welcome break to the respondent, giving them an active task to break up the long sequences of interviewer questioning .

The GHS pilot achieved 153 interviews. The results are displayed in Table 1.

Table 1.
Average times for self-completion sections in the GHS

	Paper (concurrent)	**CASI (sequential)**	**Excess of CASI(sequential) over Paper (concurrent)**
Family Information	2 minutes	5 minutes	+3 minutes (150%)
Contraception	2 minutes	6 minutes	+4 minutes (200%)
Drinking	4 minutes	7 minutes	+3 minutes (75%)
Smoking	1 minute	4 minutes	+3 minutes (400%)

Obviously the time spent by household in each of these sections depends very largely on the size of the household and how much they have to say for each section. For example, a household where an individual smokes 20 cigarettes a day will take longer to complete the form than a household of non-smokers.

Due to the nature of CASI, each respondent must enter information sequentially on to the laptop. When the sections are conducted via paper, all members of the household can enter information

concurrently. This has a large bearing on the increase in time for respondents to complete these sections. It also means that the two methods of data collection are not directly comparable.

It can be seen that, on average, it took an extra 3 minutes to conduct each section by CASI rather than on paper. However the relative increases are large, as much as 400% for smoking. Over the length of the entire GHS pilot interview, CASI added 13 minutes to the length, taking the mean time from 80 minutes to 93 minutes. This would be a large increase in respondent burden if implemented on the production survey.

ATLAS was used to calculate the times in the Blaise Audit Trail from the instrument. This information was used when drawing conclusions from the pilot study. With the help of ATLAS these times were able to be measured very quickly and in a form which enabled research staff to run more in-depth analysis than would have been possible without this package. Times were calculated both for sets of questions and individual questions.

One issue that the analysis investigated was if there was any time saving in using computer-assisted coding (CAC) in Blaise for large coding frames rather than coding from a reference manual. We were particularly interested in occupation coding, which our interviewers code at home after the interview. We did not have a direct comparison of the two methods for the same coding frame, so we explored the issue by an indirect argument. The Audit Trail told us that it took almost exactly the same mean time for interviewers to code occupation by CAC (their normal method) as to code industry from a reference manual (also their normal method): 52 seconds and 53 seconds respectively. However, the cognitive task involved in coding occupation from a reference manual is known to be much greater than the cognitive task in coding industry: occupation is coded to 5 digits as against the 3 for industry, and the coding index for occupation is over 50 times longer and less structured. Therefore we were able to conclude that it would take longer to code occupation than to code industry by the same method of referring to a manual. Since coding occupation using CAC took the same time as coding industry from a manual, it follows that coding occupation using CAC is quicker than coding occupation from a reference manual.

The Audit Trail analysis also allowed us to make design decisions about CASI for the GHS. As the pilot survey showed that CASI required additional time, it was decided that not all self-completion sections of the GHS could be asked as CASI to all sets of individuals. CASI is to be reserved for those sections which are particularly sensitive and, in particular, where confidentiality from the interviewer as well as other household members might be an issue. The Smoking and Drinking sections are mainly sensitive for adolescents in the presence of their parents so, while self-completion remains mandatory for people aged 16 & 17, and optional for all other respondents, it seemed reasonable to continue to use paper rather than CASI The family information section is potentially more sensitive and for more people: it asks about former as well as present relationships and about all births/abortions and contraception. From 2003, It will be carried out in CASI. Paper self-completion documents are also available for this section for the rare occasions when a respondent is unhappy about using a laptop, or there are several eligible respondents in the household. In the latter scenario, some people can use the paper self-completion forms while others complete the section on the laptop. This will save time in the interview.

5. A pilot to measure the total time spent on new questions woven into appropriate places throughout an existing Blaise questionnaire

The European Union is planning, through its statistical office, Eurostat, a mandatory survey on income and living conditions (EU-SILC) to be carried out by all member countries. The aim of the survey is to provide information on poverty and social exclusion in the UK that can be compared with the situation in other EU countries. In the UK, the cross-sectional element of the survey will be met by adding questions to the multi-purpose General Household Survey (GHS) which already coves many of the required topics. The information for EU-SILC will be picked up from many individual questions scattered throughout the GHS instrument.

This part of the paper describes the role of the Blaise Audit Trail in the pilot work for the version of the GHS which will deliver the requirements of EU-SILC in addition to the normal GHS requirements[1]. To test the cross-sectional component of the EU-SILC in a pilot study, a CAPI instrument was prepared that integrated the GHS with the EU-SILC primary target variables that it did not already cover. Detailed timing of questions was required for this study to provide Eurostat with information on how well the questions had worked as well and, in particular, how burdensome the sections were to respondents. We needed to isolate and measure the burden that the EU-SILC questions comprised in total. This was a difficult task since the questions were in many groups scattered throughout the combined GHS/SILC pilot instrument. We decided to use the Blaise Audit Trail to provide the level of detail that was needed.

The sample design of the study involved twenty interviewers, covering the range of levels of experience expected in the live survey, to work on the EU-SILC pilot study. Probability sampling methods were used to select 20 Primary Sampling Units (PSUs). Twenty addresses were randomly sampled within the PSU and interviewers were instructed to obtain interviews at 10 addresses in order to achieve 200 household interviews. 203 individual interviews were achieved.

The EU-SILC pilot study was conducted using Blaise, like all ONS social surveys. Table 2 provides information about the length of the combined GHS and EU-SILC interview as well as the length of time taken by the EU-SILC questions alone. The combined GHS and EU-SILC timings include the interview and post-interview administration time (such as calls information and coding at home, e.g. occupation and industry). The time shown as spent on EU-SILC questions, summed by ATLAS from all over the questionnaire, represents only time actually spent in the interview.

Table 2.
Length of interview and of EU–SILC module

Pilot study elements	Household Questionnaire	Individual Questionnaire	Complete interview
GHS & EU-SILC*	17 minutes*	1 hour and 48 minutes*	2 hours and 5 minutes*
EU-SILC	4 minutes	42 minutes	46 minutes

*Including administration.

[1] The GHS is published each year as *Living in Britain. Living in Britain 2001* is a web publication at http://nswebcopy/lib2001/index.html

Table 2 shows that the EU-SILC components (whether already covered in the GHS or covered by additional questions) accounted for an average of 46 minutes of the total interview length in the pilot study. However, it should be noted that because (as we knew from other data from the study) respondents had so much trouble understanding the EU-SILC household income questions[1] the interviewers did not feel able to do what they normally do on ONS income surveys and encourage respondents to find documentation. Hence the questions did not take as long as they would have done if implemented in a production survey. As a result, the final report to Eurostat stressed that these data underestimated the true burden of the EU-SILC components and recommended changes necessary to reducing it. It also stressed that the GHS component was over-estimated in these results, partly through consequential effects of poor EU-SILC questions and partly through the inclusion of administration time. These findings point to the importance of careful interpretation of the results. The Audit Trail and ATLAS can provide measurement but not analysis.

Detailed timing of the entire interview and of the selected areas of interest was provided by the Audit Trail. ATLAS enabled the research team to easily divide the EU-SILC module from the rest of the questionnaire.

The ATLAS tool proved invaluable when analysing data over a number of cases with the way it aggregates timings. In this manner we were able to look very quickly at the length of the interview and make rapid changes where necessary. Once again, the data in this readable form was able to be transferred into SPSS where further investigation could be conducted. As mentioned above, it was important to look at specific sections and how each section related to each other. This can give an indication of whether particular sections are influenced by other parts of the questionnaire.

6. Conclusions

The Audit Trail facility has become a more usable and adaptable tool which can now be used for analysis. When we first began to use the Audit Trail there was a fear that the amount of data produced would have a negative effect on the performance of the laptop. There has been no evidence of this. We are, therefore, continuing to use this function where detailed timing information will prove useful when analysing data or there is a pressing need from the client for detailed times in a questionnaire.

ATLAS is a very useful analysis tool but further development of the software (and other systems if necessary) would be required to fit ONS requirements for routine/automatic use.

These developments are:

- ability to distinguish between different types of case (e.g. responding, non-responding cases);
- outputs should have min, max and range of values (field, block, subsections, datamodel) as well as averages;
- data should be able to feed into other information management systems (e.g. response rate, interviewer performance monitoring both for individually and divisionally;
- the user interface could be improved, so that you can look at different reports at the same time (e.g. different months of same survey).

Despite our wish to see these desirable additions, ATLAS has already greatly enhanced our ability to utilise the Audit Trail facility and we will continue to use it for future analysis. We are very grateful to Statistics Canada for permission to use it. Audit Trails have proved useful in both the timing and

[1] Not designed by ONS!

testing of questionnaires as well as trying new procedures and ways of carrying out data collection. It has been interesting to see what sections take the interviewer a particularly long time to get thorough and to assess where, if any, changes can be made to make interviews less burdensome to the interviewers and respondents.

As the Audit Trail has, as yet, had no visible adverse effects on data collection, there is the ability to collect information routinely on all SSD surveys. With the help of ATLAS this data can now be quickly transformed into a readable form for analysis. The fact that data can be quickly extracted has led to its increased use on designing surveys and questions. We have recently used it to aid the design of the *People, Families and Communities Survey* which is intended to find out about the role individuals and families play in their local community and to explore issues related to social capital.

The Audit Trail has been used on the pilot of this survey to give detailed timing information to the research team. From this information they were able to provide feedback to clients about if there were any particularly long sections or any subject matter where it took more time than expected to collect data.

There is also the potential to use Audit Trail data to give an indication of interviewer performance. For example, it is possible to compare the times interviewers take to complete individual sections of an interview. with due attention to our previous warnings about the need for careful interpretation, this method can be used to judge an interviewer's effectiveness. Audit Trail data can be used more straightforwardly for probity checking. In the future, we would like the audit trail information to aid in monitoring interviewer performance and also as a training tool.

We would also like to take further advantage of some of the other features of ATLAS such as looking at the number of fields which were answered *don't know* and *refusal*, and at changes or edits to questions. It will be interesting to find out if there is a certain type of question which often has its value changed or edited.

We are continuing to increase our use of the Audit Trail facility. It has become a standard tool for the design of surveys. It is implemented on nearly all ONS social Surveys as it has no visible affects to laptop performance or at any stage of data collection – including getting information into the office.

References

Bumpstead, R., (2001) "A practical application of Audit Trails", 7th International Blaise Users Conference

Hansen, S. and Marvin T., (2001) "Reporting on item times and key strokes from Blaise Audit Trails", 7th International Blaise Users Conference

About the Author

Tim Burrell is a Senior Social Survey Researcher at the Office for National Statistics. He is currently working on the Expenditure and Food Survey (EFS) as well as a member of the Blaise Development, Standards and Support Team (BDSS) within ONS. In over 5 years at ONS, he has worked on a wide range of surveys including a survey of adult dental Health, sports and disability and a survey of drug offending in communal establishments. He can be contacted at ONS, 1 Drummond Gate, Pimlico, London SW1V 2QQ tel. 0207 533 5510; fax 0207 533 5100; e-mail tim.burrell@ons.gov.uk.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

The Survey Process in the Digital Age. Factors Limiting Transfer of Technology and Know-How. A European Research Perspective

Karen Barrie

Abstract

The digital revolution and the increased value accorded information have elevated expectations regarding the timeliness, relevance, reliability, comparability and scope of statistics. This is particularly true within the context of the European Union, where exacting requisitions are made of all countries, regardless of the efficacy of their statistical systems. At the same time, there is a constant pressure to reduce data collection costs, lighten the burden on data providers and protect confidentiality. Such demands require traditional statistical concepts, measures and technologies to adapt continuously in order to retain their pertinence and effectiveness for users.

A proliferation of new technologies promises to redefine the survey process, raising implications for sampling, data collection, analysis and dissemination, whilst enhancing the possibilities for secondary and comparative data usage. Recognising this, considerable research and development effort has been expended within this problem space, often yielding potentially far-reaching results. However, realising the underlying expectations presents a number of challenges.

For many years, key stakeholders within the European Statistical System have been concerned about the poor take-up of the results of R&D in statistics, notably technological innovations, together with the limited transfer of technology and associated know-how amongst statistical agencies. Following the under-utilisation of technological outputs from both the 3rd and 4th Framework Projects, AMRADS (an Accompanying Measure to Research And Development in Statistics) was initiated in a bid to prevent 5th Framework products suffering a similar fate. In addition to strengthening the conditions for ensuring concrete and tangible technology transfer, AMRADS sought to ascertain why some NSIs are good at adopting new products and ideas, whilst others are less assimilative.

As the 5th Framework draws to a close, this paper explores the reasons for the apparent disconnect between the promise of emergent technologies and actual applications to the survey process. The experiences of two specific European R&D projects are considered, armed with insight secured via AMRADS and recommendations to address the emergent issues are made.

Technological advancement presents a clear means of addressing the apparently conflicting demands made of survey producers, but in turn introduces a series of organisational, cultural and indeed further technological concerns. It is vital that these inherent issues are understood and characterised if future innovations are to be fully exploited and ultimately, the information needs of policy-makers, businesses and the citizen are not to be under-serviced.

1. Introduction

A proliferation of new technologies hold the potential to redefine the survey process and, as numerous contributions to this conference attest, many exciting research and development (R&D) initiatives are currently being pursued to this end. However, the take-up of recent R&D outputs across Europe has been rather disappointing and the transfer of technology and know-how (TTK) amongst statistical agencies remains somewhat fragmented. The EU Fifth Framework Programme entitled *an Accompanying Measure for Research and Development in Statistics (AMRADS)* was designed to address these concerns.

This paper firstly reviews the reasons why TTK has gained elevated importance within the context of the survey process, before describing the context for the AMRADS initiative more fully. The *AMRADS* findings to date are then outlined, together with emergent statistical agency take-up issues. The importance of a closer coupling between technology transfer and collaboration is then considered. Finally, two recent R&D project experiences are evaluated and factors critical to successful transfer in this collaborative environment identified, suggesting possible ways forward in future.

2. The Need for Transfer Across the European Statistical System

The digital revolution and the increased value of information have elevated expectations regarding the timeliness, relevance, reliability, comparability and scope of statistics. This is particularly true within the context of the European Statistical System (ESS), where Monetary Union, the Stability Pact, Enlargement, etc, have placed exacting requisitions on all countries, regardless of the efficacy of their statistical systems. At the same time, there is a constant pressure on the ESS to reduce data collection costs, lighten the burden on data providers, offer greater flexibility and protect confidentiality. Such demands require traditional statistical concepts, measures and technologies to adapt continuously in order to retain their pertinence and effectiveness. Moreover, the digital government movement has brought issues of transparency and accountability to the fore. Statistical data and products now constitute a key part of national and international knowledge bases and satisfying the information needs of diverse users has become a major priority, competing for attention with the traditional focus on processing efficiency. Consequently, the visionary elements of the Statistical Information System can now be mapped to two dimensions:

- Participation in a consistent and intelligible ‘global’ statistical user environment
- Efficient, respondent-centric, metadata-driven processing

Technological advances present ever-new opportunities to realise aspects of this vision and the survey process needs to be constantly analysed and refined to take full advantage of such dynamic developments. However, such an undertaking is costly. Most statistical agencies are facing severe resources constraints and often do not possess the funds or the expertise to carry out all the work that they would like. When faced with economic pressures, it is often the agency R&D budget that suffers. As agencies across Europe typically face similar issues and policy challenges, there is much to be gained from pooling and sharing know-how in order to reduce re-invention and solve common problems. Whilst a strong tradition of co-operation exists within some statistical agencies, membership has remained rather limited.

Moreover, whilst inter-agency exchanges can provide an efficient and effective means of finding answers to near-term problems, there is an additional need to be both forward looking and outward looking. Indeed, as one statistical agency has conceded, the tradition of collaboration amongst

agencies, although positive, has "too often resulted in introverted solutions and missed 'technology-led' opportunities". (Joshua, 2003:2). The academic environment supports the exploration, exploitation and further development of emergent technologies in a way that statistical agencies often cannot afford. Funded research that considers potential applications for data collection, access and integration technologies, together with innovative simulation and presentation techniques all appear to be of immense importance to statistical agencies. However, take-up of the results of such R&D efforts has to date been rather poor.

3. Technology and Know-How Transfer Concerns: The Context for AMRADS

For many years key stakeholders within the European Statistical System have been concerned about both the disappointing take-up of the results of R&D in statistics and the limited TTK amongst agencies. These concerns have applied to research conducted under the Information Society Technologies (IST) programmes and under national as well as other initiatives. Specifically, the outputs from SUP.COM, of DOSES (Development of Statistical Expert Systems) under FP3 and DOSIS (Development of Statistical Information Systems) under FP4 have been under-exploited in the statistical production process. Those agencies with a real need for imports of more advanced statistical methodologies and associated technologies have experienced little benefit from the research.

These concerns were reflected in the decision to promote TTK taken by the EU Statistical Programme Committee as long ago as September 1997. EuroStat, in conjunction with the Joint Research Centre in Ispra, took certain steps, notably setting up the European Statistical Laboratory (ESL) and convening a TTK conference in Prague in October 1999. However, although these steps were commendable, practical progress has been rather limited and slow. The ESL has so far had more of a virtual rather than a real existence and the results of the Prague meeting were somewhat different than expected. The danger that the same degree of under-utilisation of research outputs and poor TTK could occur under the Fifth Framework was recognised and the need to launch systematic remedial effort within the ESS identified. *AMRADS* was devised as part of this remedial effort.

4. The AMRADS Experience

As discussed, *AMRADS* seeks to support transfer and co-operation between agencies and to strengthen the conditions for the take-up of the results of R&D efforts. The project began by attempting to better understand agency needs for methods, technology and know-how and to determine the most appropriate mechanisms for transfer. Questionnaire analysis resulted in the establishment of six thematic groups, a training event being scheduled for each, informed by activities intended to establish more precisely the needs within each theme, existing best practice and the feasibility of transfer.

At the time of writing, only two training workshops had been conducted. However, the informative actions had secured an understanding of the best methods for transfer within specific contexts plus some insight as to why certain statistical agencies have been better at adopting new products and ideas than others. Patterns regarding the inclination for certain countries to share expertise also emerged, with the Nordic countries having assumed a particularly proactive stance in this respect. Specifically, the activities found that there is a general willingness to engage in inter-agency exchange and collaboration and that forums serving to bring agencies together to share experiences are conducive to the extension of existing networks, sewing the seeds for future inter-agency collaborative developments. Whilst know-how and methods have generally been well received, the goal of helping

to create the conditions for take-up of R&D technology products however required deeper investigation. The following section considers some of the barriers yet to be conquered.

5. Agency Barriers to Technology Take-Up

It is often perceived that lack of agency awareness of R&D activities constitutes a major barrier to technology take-up, aggravated by inadequate access to consolidated and appropriate information. However, even where such factors are not problematic, it appears that many statistical organisations have a poor perception of academic R&D outputs and a misunderstanding of their potential applications.

Further, as variable assimilative behaviours have been observed across agencies, agency culture also constitutes a key determinant. Here, pride of ownership can play a major role in deterring the adoption of R&D outputs; the perception being that internally generated technologies are far superior to whatever could be created in an academic setting. Such reluctance may be compounded by the belief that the agency is too different and that products developed elsewhere will not translate. Indeed, even agencies that have demonstrated a willingness to adopt external know-how, methods and technologies may still attach significant importance to the in-house development of solutions as the only way that permanent development and progress of employees and the institution can be supported. (Katic, Klanjscek & Kozjek, 2003).

The context also determines, or at least influences, the type and extent of technical solutions to be embraced. Agencies now have to deal with a complex infrastructure and complex statistical processes. This complexity demands careful planning and the gradual introduction of new processes and integral infrastructure solutions. Timing of transfer approaches may thus prove crucial. New systems have to be integrated smoothly and completely in the survey environment, taking account of all methods offered to respondents and all steps in the production process. (LaBillois, 2003). This extends beyond far technological operability requirements and can require a deep understanding of the current environment. Awareness of similarities in circumstances and experiences may partially explain the greater tendency to adopt solutions developed by other agencies, rather than the products of academic R&D projects. More significantly however, issues of security and service reliability mean that agencies must assume a very low risk tolerance, demanding the adoption of industry-proven technologies and solutions. Further, cost-effectiveness concerns can restrict acquisition or development of solutions to a limited range of technologies to reduce the cost of ownership.

Together, these issues point to the clear need for greater agency involvement in research projects and the requirement to enhance mutual understanding and awareness through ongoing communication and feedback loops, thus underscoring the importance of a closer coupling between collaboration and technology transfer.

6. Technology Transfer and Collaboration: A Closer Coupling

It has long been recognised that the traditional model of technology transfer, in which researchers work autonomously and then disseminate their findings is not appropriate within the statistical environment. Where the research objective is to develop a product to be used in statistical agency activities, there is a general agreement that extensive co-operation between interested statistical agencies and the academic researchers will be required. Moreover, where academic researchers have

an idea that they believe could eventually be developed into a more widely marketable product, it is commonly believed that the options for achieving this will be very limited unless the commercial sector is involved from the outset. However, despite this consensus, the media, mechanisms and motivations for such participation remain less clear and there is still much to be learned about satisfactory approaches to technology transfer and collaboration. To illustrate the problems encountered during both types of collaboration, the experiences of two recent R&D projects are considered.

7. Technology Transfer and Collaboration: Two Recent R&D Project Experiences

The first project sought to provide a modular system of software that would enable providers of statistics to publish data in a unified and unifying framework and allow consumers to access it in an informed manner with minimum effort. Research outputs included know-how, models and a software prototype, for which a number of success criteria were determined. At the project's inception it was believed that outputs offered high potential for exploitation by data providers such as statistical agencies and also third party hosts, notably data libraries. The project consortium comprised a mix of academic and statistical institutions.

Efforts sought to build upon an earlier initiative, from which better understanding of the difficulties that target organisations have in integrating new technologies had been secured. Recognising that statistical agencies were already investing heavily in the supply of data via the web, the research sought to concentrate on truly innovative areas where value could be added. The project was very much led by academia and although agencies were involved, their role was largely reduced to informing initial requirements and prototype evaluation.

The result was a solution to a perceived problem, acknowledged as of potential future interest to agencies and data libraries, but perhaps not a current priority. The focus on value-added features meant that more fundamental requirements were neglected, impacting usability. Moreover, although some insight into the operational constraints of agencies had previously been secured, the implications of the need to balance the open culture of the Internet with the closed ethical code that agencies are obliged to follow had not been fully understood. Consequently, although favourably evaluated, the amount of additional development work needed to make the software product usable in a given context remained quite considerable, reducing the likelihood of take-up in its current form.

The second project concerned data collection technologies and in addition to the development of know-how, models and software also sought to contribute to metadata standards. At the time of writing the project was ongoing and a number of issues regarding transfer possibilities had still to be fully resolved. As such, only a few comments are made in this respect.

At the project's inception, it was believed that the generic solution to be developed would be of interest to a variety of market sectors, suggesting a requirement for involvement from the commercial sector. Whilst finding a commercial partner often presents the greatest problem, in this instance a suitable enterprise was identified, bringing considerable commercial software development experience and relevant market knowledge to the project consortium.

Difficulties in attempting to merge the business and academic worlds are commonplace due to the very different orientations of the two groups. In this instance, faced with limited project resources, these manifested in the form of opinion clashes as to which aspects of the software 'prototype' attention should be devoted to. In particular, compromises regarding the inclusion of features deemed

essential to secure market interest, but already present in commercially available products, perceived by academics as detracting from the time available for more exploratory and innovative developments were hard fought. Intellectual property rights also required very careful handling and were the subject of a number of misconceptions amongst the academic partners.

These two experiences indicate that there are a number of barriers that must still be overcome in successfully transferring technology from the research environment to the survey process.

8. Barriers to Transfer in the Collaborative Environment

A major transfer issue emerging from the described experiences is the lack of a common understanding of the complexity of the innovation process, the purpose and scope of research prototypes, resulting in mismatched stakeholder expectations as to what might realistically be achieved within the context of a given research project. Much of this confusion may be attributable to the blurring of boundaries between basic and applied research.

Whilst basic and applied research were traditionally sees as activities of quite different natures, in the 1980's information technologies started a trend whereby it became increasingly important to turn scientific research into marketable products, placing increased emphasis on the relevance of research. (EMBO, 2002). Consequently, the way in which researchers co-operate and forge partnerships with industry and commerce changed dramatically throughout the 19990's. (Wirth, 1995). Today, interest in technology transfer has blossomed into a major visible element of economic policy. However, the path towards relevance-oriented research is not free of pitfalls. Faced with economic change, stagnating government research support and various TTK funding incentives and stipulations, there may be a tendency if not to deliberately oversell the TTK potential of research ideas, then to understate or underestimate the amount of residual effort required to progress from pre-product to fully functional tool.

Whilst most technology emerging from academic research is embryonic, statistical agencies need readily implementable solutions that fit their environments. Formal targets and human nature dictate that marked changes must be seen within a short timetable. Strategies are still evolving and will continue to do so for as long as the technology is developing very quickly and as the perceptions and habits of users are changing over time. A solution that is developed today might need to be replaced by something significantly different in a relatively short time. (LaBillois, 2003). In this climate, the lengthy post-research development times and further investment needed to implement R&D outputs are undesirable and, where unanticipated, will inevitably be met with disappointment.

Commercialisation represents an alternative exploitation pathway. Here the ability of a research institute is often hampered by the fact that the primary mission has historically emphasised research, publication and teaching, not the development of products, precluding the investment of time and resources to adequately address the demands of transfer activities. Whilst the importance of technology transfer in the new economic environment has now been recognised, academics are not best qualified to carry out the various market research, marketing and business management functions required, nor arguably should they be.

Commercial partnership introduces another set of challenges, as the gap between the missions of commercial organisations and academia is vast. The focus of universities has traditionally been to support long-term research programs that support faculty members over several years. This conflicts with the objectives of many co-operative development programs, where fast results are required. If a

clear understanding of research goals, priorities and roles is not secured, this may render lines of communication problematic. The basic bureaucratic processes related to TTK can also be formidable. Moreover, the intrusion of private enterprises and funding into the academic world is not to everybody's liking and can create conflicts of interest, notably where responsibility for policy of protection and also dissemination rests with one individual. Resolution of such conflicts, together with negotiations over legal issues, potential value and exclusive rights can introduce significant delays.

9. The Way Forward: Supporting TTK and Collaboration Within the ESS

So what can be done to support the dual activities of TTK and collaboration within the ESS in future?

If we look to the USA, here the National Science Foundation's (NSFs) Digital Government Initiative has already successfully nurtured and supported partnerships between statistical agencies and academia and has identified a number of steps to promoting statistical agency-research community interaction. (Hert, 2002). One particular workshop dedicated to exploring means of further enhancing interaction levels uncovered a number of preconditions that appear equally applicable within the context of the ESS. Specifically, the importance of appropriate funding for technology transfer and collaboration was stressed, together with the need to understanding and clarify the various roles of industry, researchers and agencies.

When contemplating possible ways forward, a series of recommendations can be made. The first concerns the development of the research proposal and the importance of ensuring that comprehensive TTK plans are not only present, but also are consistent with the overall research aims, the level of agency support (or commercial viability) and funding provision. Proposal evaluation criteria will similarly require adaptation to confirm stated levels of interaction and support and to verify their compatibility with the project TTK goals.

In addition to the need for role clarification, the discussion of the two European R&D experiences highlighted the importance of ensuring shared expectations amongst all stakeholders. To this end, the scope and nature of the research 'prototype' must be clearly defined, priorities agreed and the remaining effort required rendered readily discernible.

Where the intention is to use research results directly in statistical agency activities, a key theme to emerge from the NSF workshop was the need not only for early agency involvement, but also for interaction throughout the project lifecycle. The EU R&D experiences discussed indicate that such involvement is particularly pertinent during the design stages. Moreover, the agency must be accepting of the result. Attention must also be paid to usability issues and operational constraints within the given context. This may constitute a point of departure for some researchers, necessitating close working with agency staff. In turn, the agency must ensure awareness of the project's aims and progress throughout parts of the organisation to be affected by its transfer. Expectations regarding supporting material must also be clearly set. In particular, experience has shown that researchers may constitute the best choice for user manual and guidelines production, even where provision is made for such deliverables within the research budget.

The actions suggested above however all presuppose a willingness on the part of statistical agencies to participate in collaborative R&D initiatives. As discussed, when faced with budget cuts and concerns about the ever-changing technological landscape, many agencies seek solutions that can be more readily implemented. The merits of participation in R&D efforts must be made clear. To this end, the research community needs to become more proactive and better able to market the potential benefits

and uses that are enabled by new paradigms and technologies. The fruits of previous successful partnerships must be effectively showcased if agency participation in TTK is to be enthusiastically supported, rather than perceived as a distraction from the real job. The importance of the role of technology transfer as a specific and permanent function within the statistical agency has also been underlined, demanding adequate resource provision, an appropriate organisational model and strong support from top management. (Masselli, 2002).

Agency desire to participate in collaborative R&D projects is likely to increase where the research is of a more applied nature. In such instances however, academic researchers may face barriers to participation. Importantly, reward systems and research assessment activities must change if researchers are not to be penalised for their involvement.

The willingness to work together is also likely to increase where trust is established and mutual understanding and respect for each other's worlds secured. As Hert (2002) asserts, collaboration is critical not only to technology transfer, but also to efforts to build a robust community. For this to happen, more venues such as this conference that bring together research and industry are needed to support mutual exchange of ideas and needs. Maintaining collaborations in the long-term is also critical, highlighting the need for dedicated funding. Moreover, new and varied exchange models are required. Staff mobility is a much under-utilised transfer vehicle within the statistical domain.

Where exploitation through commercial engagement is sought, a significant barrier is often the identification of a suitable partner to provide the requisite business and marketing skills. Here the establishment of a suitable matching service could prove beneficial. Where partnership is established, project objectives must again be clearly defined, priorities agreed and intellectual property issues fully addressed in the first instance.

Whilst computer science may once have been guilty of conducting basic research in isolation from the real world, it is vital that the pendulum does not swing too far. TTK must not be allowed to drive the academic research agenda and the dualism between basic and applied research must be preserved. Where proposed research activities are more basic in nature, a distinction should be drawn between pure research, which must remain unconstrained by short-term agency needs, and more problem-driven research, which explores issues of known interest to agencies and seeks to produce new technologies of strategic importance. In the case of the latter, TTK plans are needed, but these must be appropriate. Such research is quite different from truly applied research and immediate exploitation or commercialisation through industry collaboration and entrepreneurship may not be practicable. Instead, a bridging mechanism may be required.

Research on the technological innovation process per se highlights the role of intermediaries. (Dodgson & Bessant, 1996). As the name suggests, intermediaries operate between users and suppliers of technology and help to create the links within networks and systems by bridging between institutions, encouraging interaction and assisting with undertaking search, evaluation and dissemination tasks. They ensure that technological know-how is broadly dispersed within the system and can provide a compensating mechanism for weaknesses or "holes" in the system. This role is gaining increased governmental recognition, as evidenced by various funding initiatives such as the Proof of Concept fund, together with the development of a number of Intermediary Technology Institutes. However, the extent to which survey-oriented research stands to benefit from such efforts is questionable.

10. Conclusion

Many positive short-term developments are currently taking place inside the statistical agencies that comprise the ESS and the inter-agency collaborative network is widening. At the same time, a number of promising ideas are being explored through research. However, the transfer of the results of such endeavours remains weak. This paper has considered the issues inherent in various transfer mechanisms and whilst several recommendations have been made, numerous challenges remain. If TTK is not to constitute the Achilles' heel of the European Statistical System, a new balance must be struck. Only then will the survey process be able to truly capitalise on the abundant opportunities of the digital age.

References

Dodgson, M. and Bessant, J. (1996). *Effective Innovation Policy: A New Approach*, London: International Thomson Business Press.

EMBO: European Molecular Biology Organization. (2002). Balancing basic and applied research. *EMBO Reports* 31(11).

Hert, C. (2002). *Challenges in Statistical Digital Government: Promoting Statistical Agency-Research Community Interaction: Workshop Report.* Retrieved 21st April 2003 from http://www.digitalgovernment.org/

Joshua, D. (2003). *A New Dawn for Information Technologists*. Retrieved 19th April 2003 from: http://www.unece.org/stats/documents/ces/ac.71/2003/18.e.pdf

Katic, M., Klanjscek, M. and Kozjek, P. (2003). *Management and Monitoring of Statistical Process and Impact on Data Quality*. Presented at the 1st Project Meeting of CODACMOS, Rome, March 2003.

LaBillois, T. (2003). *The integration of Electronic Data Reporting (EDR) and other new systems in the ongoing collection process*. Presented at the 1st Project Meeting of CODACMOS, Rome, March 2003.

Masselli, M. (2002). *Transferring Technology in Data Collection.* Retrieved 11th April 2003 from: http://webfarm.jrc.cec.eu.int/etk-ntts/ Papers/final_papers/11.pdf

Wirth, N. (1995). *Effective Technology Transfer.* Retrieved 18th April 2003 from: http://www.oberon.ch/services/odf/tribune/trib1/efftech.html

About the Author

Karen Barrie works at the Centre for Educational Sociology of the University of Edinburgh. She may be contacted at: Centre for Educational Sociology, Moray House School of Education, University of Edinburgh, St. John's Land, Holyrood Road, Edinburgh EH8 8AS, e-mail: karen.barrie@ed.ac.uk.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Software Process Improvement Efforts at the US Census Bureau. Experience of the Economic Directorate

Howard Hogan & Ellen Soper

Abstract

This paper examines the plan to improve the way statistical software is developed to support surveys at the US Census Bureau. It examines the existing culture and the need to bridge the gap between management expertise and software programming expertise required to manage organizational and technological change. It is the goal of the Economic Directorate to evolve toward a methodology that integrates software engineering and management excellence with survey research and computing. The paper discusses the general approach, including the management steps necessary to make the transition toward software process improvement. This paper looks at challenges such as the limited management and technical resources available, and areas of progress. In order to improve the way software is developed and managed, we need to recognize the factors that affect change and implement strategies to ensure successful implementation.

Keywords

Capability Maturity Model; Software Process Improvement; Economic Programs; Current Surveys

1. Introduction

With the introduction of UNIVAC I in 1951, the US Census Bureau became the first statistical agency in the world to enter the "computer age." The organizational structure has passed a legacy that continues to today. In the early years most software programming was done by mathematical and survey statisticians. Standards for programming and documentation were rare resulting in few "best practices" to propagate. On the other hand, communication was great; the person who developed the software was the same person who analyzed the data produced by the software. The analysts, considered the end user of the software, knew what they wanted the software to produce quite clearly. As programs became more complex, a separation grew between programmers and analysts. The programmers became more disciplined in software engineering by demanding written requirements and more formal management structures.

In early 2000, the Census Bureau established the goal of improving the overall software process by implementing some of the key practices of the Capability Maturity Model for Software, commonly called, CMM (See below, Section 3). Understanding the current culture, determining what new behavioural changes are needed and setting the course for successful change is a difficult endeavour. The Census Bureau recognizes the need to change the existing culture in order to bridge the gap

between business expertise and software process expertise required to manage organizational and technological change.

A management group in the Economic Directorate was established to manage a software process improvement initiative by following the principles of CMM. The group began by developing their own program plan, schedule, change control mechanisms, and success criteria. This planning of the planning (which we refer to, somewhat in jest, as Meta planning) and the lessons learned in implementing CMM will form the heart of this paper.

2. Existing Culture

The Census Bureau has an established culture in terms of how software development and survey processing is accomplished. In this culture, statisticians document the software requirements as software specifications. Software specifications are not strictly user requirements. They are closer to, but not quite, what the software industry might consider the result of structured analysis and design. They tend to be written in the language of equations, algorithms, and logic tables. Since statistical analysts write these documents, tools common in the software industry, such as an Entity Relationship Diagrams or Data Flow Diagrams, are seldom used. The software specifications are quite technical and few people are able to ascertain the true requirements. The specifications do not serve as an effective way of communicating the requirements of the system with others in the organization. This presents a problem when a statistical analyst or programmer leaves the project, and someone else takes over. Communicating the software requirements to a wide community of system users is also a challenge.

Another aspect of the culture is that the statistical analysts are accustomed to being able to make frequent “improvements” to the system. Myriad informal arrangements have been built up over the years to work around the formal specification-code-test process. In addition, statistical analysts often use the files that are output from the programs; they need to know the database structure so they can write program scripts and conduct further data analysis. They have a need to know how files are structured, organized and related, and how the data are processed. On the other hand, programmers are also highly knowledgeable of survey processes. They may have worked on processing the same survey for years. A good aspect of this culture is the frequent and successful communication between the analysts and the developers. The downside of this culture of direct interaction is that establishing formal management control mechanisms, such as configuration change control, can be a major challenge, especially when developing large, generalized software systems. Few of the mid-level managers developed project schedules or assessed project risks. Requirements were poorly defined and there was little coordination of resources and how they were allocated.

Since projects were being accomplished successfully meeting critical milestones, there was little doubt the Census Bureau had skilled analysts, programmers and managers. However, over time we relied more and more on software to achieve goals. Through technological advances to meet more complex demands, we realized a need for a new, more structured framework from which to manage the survey work. The time was right to embrace change and begin an effort to improve the way business was accomplished, specifically within the Economic Directorate of the Census Bureau.

3. CMM Overview

The CMM, developed by the Software Engineering Institute, is a framework that describes the key elements of a software process (Carnegie Mellon University, 1995, Humphrey 1989). It covers practices for planning, engineering, and managing software development and maintenance activities, and employs an evolutionary improvement path from an *ad hoc*, immature process to a mature, disciplined one. It is the goal of the Census Bureau to evolve toward a culture that integrates software engineering and management excellence with survey research and computing. As depicted in Figure 1, CMM has five levels from Level 1: Initial to Level 5: Optimizing. The Census Bureau's focus is on the key processes of Level 2: The Repeatable Process.

> *By the time an organization reaches Level 2, its software process is repeatable and under basic management control. Project managers are able to make reasonable estimates and project plans, and to track and control project performance via these plans fairly consistently. [Kenneth M. Dymond, "A guide to the CMM", page 2-1]*

CMM-Level 2 includes six key processes:

- Software project planning
- Software project tracking and oversight
- Software configuration management
- Requirements management
- Software quality assurance
- Software subcontract management

The CMM is a proven framework, which can significantly improve the software development process and management of software development and maintenance efforts. To achieve a certain level of the CMM you must meet the goals of each of the key processes at and below that level. The CMM is not a particular lifecycle model for software development, such as the "water fall" or "V," nor does it rely upon a particular set of software tools. The CMM describes the practices of an effective software process, but the organization determines how the practices for a successful project will be executed.

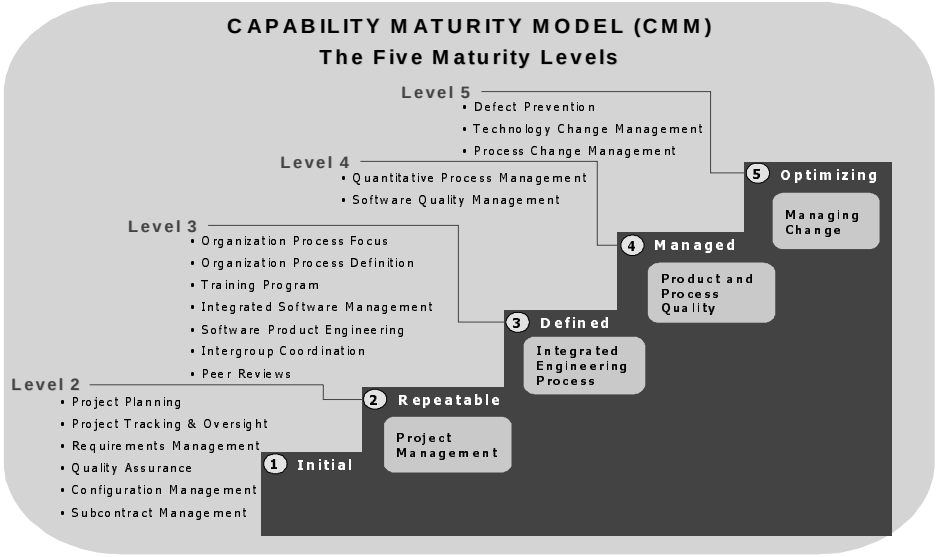


Figure 1

4. Laying the Groundwork: CMM Pilot Program

The Census Bureau wanted to investigate the possibility of implementing the CMM on a corporate level. It conducted a pilot that included three different projects from across the organization to see if the CMM framework would work in our environment and culture. The overall goals of the projects were to improve their software development processes by implementing CMM practices such that they would achieve most of the CMM Level 2 compliance activities.

The Economic Directorate sponsored one project to participate in the pilot. The CMM Pilot Program, as documented in the Census Bureau's CMM Pilot Program Final Evaluation, had a positive effect on each of the three pilot projects. It demonstrated that the CMM model (at least at Level 2) could work at the Census Bureau. The approach gained much publicity and acceptance as evidenced by the feedback received from the pilot participants who indicated a desire to continue using the CMM to improve their software development efforts. Although none of the three pilots achieved Level 2, all three pilot projects expanded their original scope and incorporated software process improvement efforts of their own accord. Judging by the increasing implementation of software process improvement activities across the Census Bureau, there is a growing perception that software process improvement is a worthwhile investment.

The results of the pilot program included many lessons that were learned and further recommended for areas interested in investing in software process improvement. One important lesson is that implementing the CMM is not a cookbook exercise. Software process improvement activities must be targeted toward basic business practices to be effective. The biggest improvements were realized when the focus was on the pilot projects' needs - not just trying to implement each activity as prescribed by the model. The implementation of CMM involves investigating how projects currently do work and evaluating specific activities to improve the problem areas. The CMM can be used most effectively as a resource for suggested improvements.

Another lesson is that grassroots support is key, but equally as important is the need for strong upper management support and backing of the CMM principles. Implementing CMM involves upfront investment and time to be accepted by all. Training is also another important aspect of implementing the model. All of the project participants obtained at least an introductory course of the CMM. However, even with this training, project participants didn't fully understand the benefits of the model until they tried to implement portions of the model.

Other lessons include the realization that implementing the CMM is not an easy task; it requires more resources than ordinarily available to project managers. The pilot program had additional resources to support the effort. Specifically, each of the three projects had a part-time project coordinator, documentation specialist, quality assurance manager, and CMM expert consultant over and above what is normally assigned to Census Bureau projects. By participating in the pilot, we learned we can't implement an overall process improvement effort without considering adding more resources, providing opportunities for training, or considering a smaller scaled, incremental approach that can be folded into normal project work.

5. The Experience of the Economic Directorate

The Environment

With the completion of the pilot program, the Economic Directorate was committed to expanding the Census-wide improvement efforts to some of their major economic programs to help build a new management framework designed to target many of the issues that have existed for years.

Five divisions form the infrastructure required to manage the economic programs.[1] There are three subject matter divisions: Service Sector Statistics Division, Manufacturing and Construction Division and Company Statistics Division. These divisions are further divided into functional areas largely comprised of subject matter experts, referred to as analysts, who plan the work such as developing the survey questionnaire and developing the tabulation requirements. In addition, they are heavily involved in reviewing, analyzing and editing the data.

In addition, there is a central coordination division, Economic Planning and Coordination Division. This division is responsible for scheduling, allocating resources, documenting requirements and other planning functions. Finally, there is the Economic Statistical Methods and Programming Division that includes both survey methodologists and software programmers.

The economic programs collect data on retail trade, wholesale trade, transportation, communications, services, manufacturing, mining, construction and finance. It is roughly divided into two major program areas, one that is responsible for the Economic Census conducted every five years and one that is responsible for the many on-going surveys, referred to in this paper as the Current Surveys. The Current Surveys include several surveys that are conducted on an annual basis. There are 11 annual surveys in all[2]. In addition, the Current Surveys include 10 important "economic indicators" listed below. With the exception of the two quarterly surveys, these surveys are conducted monthly.

- Advance Retail and Food Service Sales
- Monthly Wholesale Trade: Sales and Inventory
- Advanced Report on Durable Goods
- Manufacturing, Shipments, Inventory and Orders
- Quarterly Financial Report - Manufacturing, Mining, and Trade
- Quarterly Financial Report - Retail Trade
- Construction Spending
- New Home Sales
- Housing Starts
- Manufacturing and Trade Inventories and Sales

Those who guide economic policy closely follow the data produced by these indicator surveys. The data have an important affect on the financial markets. Given their sensitive nature, the indicator survey projects have strict scheduling requirements. Indeed, the data release is timed down to the

[1] A division has roughly 150 to 250 people, led by a Division Chief. Divisions are divided into 3 to 5 Assistant Division Chief (ADC) areas. An ADC would supervise several branches of 10 to 20 professionals. There is a total of 900 people who work in these divisions. Two additional divisions, Foreign Trade and Governments, have their own systems and are outside the scope of the project and paper.

[2] The exact number of surveys depends upon what one counts as a separate survey. For example, the Current Industrial Reports, counted here as one survey, is comprised of 57 separate components, each targeted on a different industry.

second, with the financial markets reacting within seconds afterwards. In addition, to the Current Surveys and the Economic Censuses, there are a few "periodic" surveys, which altogether define the work environment of the Economic Programs.

Historically, the functional areas within these divisions conducted much of their work within their respective area. Each area had their own programmers, statisticians, analysts, etc. Many had their own computer-processing systems that supported these surveys.

In 1997, a common processing system for all of the Current Surveys was developed. This system became known as the Standard Economic Processing System, or StEPS. The StEPS system was envisioned as an all-inclusive processing system. It is written in SAS and runs on a Unix platform. The use of SAS was considered extremely central since many of the analysts and methodologists working on the surveys had extensive SAS experience. Developing a SAS based system allowed easier access to their data to develop specialized applications. The StEPS system was designed to have standard data structures to support all aspects of survey processing, as well as integrated modules for customized functionality.

The development and migration to the StEPS system essentially created a multifaceted working environment. Initially, there was the design and development of the StEPS system. Then there was implementation of new surveys to the StEPS system, which included new functionality for those surveys. In addition, the StEPS system required on-going maintenance and troubleshooting. Lastly, there was the work involved in migrating the various current surveys from the legacy systems to the new generalized StEPS system. The migration work also included new functionality, as well as the process of transferring data and conducting testing in parallel with legacy systems. Since all systems have not yet migrated to StEPS, a certain amount of maintenance is necessary on the old legacy systems.

The Challenges

As one might imagine, working in a large, complex environment can be challenging. In an effort to define a strategy for improvement and an effective management framework, we focused our initial efforts on a manageable population of our work, the Current Surveys. The management of the Current Surveys has many problems and challenges of their own. Our first step towards software process improvement involved taking a good look at the problems we faced.

Many groups were formed to manage this work. However, clear authority and responsibility was not established. Previous to the introduction of the generalized system, the traditional "chain of command" was successful in managing the production work and legacy systems. Several other groups were established to manage the StEPS system that included a Design Architecture Review Board (DARB). The DARB focused on managing the design of the system architecture, however it evolved into a larger group. Meeting weekly, it often attracted 30 or more people. The meeting was used to demonstrate prototypes and to discuss and clarify requirements. In addition, there was a user group that had little or no management role, but was used to communicate changes to the users. There was a "StEPS Manager," although the authority and responsibility of the manger were unclear. There was no formal documentation to identify roles and responsibilities.

Changes to the StEPS system came from many sources and in many forms. They came in the form of trouble tickets, e-mails, and word of mouth. They came in the form of user needs requested through the migration process, as well as in the form of whole newly funded surveys that must be suddenly fielded. In an early effort to move to a more structured software environment, a change control board

was established. However, this board was universally seen as a failure. The principal reason given was that there was no workable system to vet the changes so the change control board spent its time on trivial changes, better handled elsewhere, and not on important issues of design, schedule and resources. In addition, there was no follow-up or enforcement of the board's decisions. We made the mistake of not clearly defining roles and responsibilities as well as, not holding those assigned to the roles accountable through consistent enforcement. One can appreciate the challenges of enforcing consistent behaviour through a cooperative arrangement with other line managers who in the past had only to worry about their own functional areas. This was the beginning of managing the work across multi-layered, organizational boundaries.

Although tasked with the same mission, many of the functional managers were not happy giving up the control they once had. Before they had their own requirements, own processing system, and their own resources to conduct their work. However, with the development of a generalized processing system, came a new way of doing business. Functional managers had to give up some of their resources to support the new StEPS system. The functional areas not only had to rely on other functional areas to develop and further enhance the system, but also had to compete with the needs of other functional areas in terms of what features to implement in the new system.

In addition to this new environment, survey migration schedules began to slip, so that surveys were left on the legacy systems far longer than originally planned. This put the surveys at risk since updates of the legacy system were put-off in anticipation of the movement to the StEPS system. Schedules were lacking in terms of this new environment. Indeed, separate schedules existed for each of the separate projects or surveys; however, far too many did not include the key dependencies that required commitments of resources from others.

A final, but very large problem, was that there was no final authority: no one to say no. There was no one to set priorities. There was no person or group to set the overall design and functionality. Those users first in line, with respect to the migration schedule, had their perceived requirements met. Those later, were simply told to wait.

Work in the new environment involved the consideration of new ways of developing requirements, new ways of interacting with others, new ways of managing resources and new ways of considering system changes. The change in technology, which seemed fairly straightforward in concept, proved to require a corresponding change in how we accomplished the work.

6. A New Framework: Meta-Planning

At the beginning of 2002, we began to systematically tackle the problems associated with the software related work within the Current Surveys. We realized we needed to help the organization to adapt new behaviours in ways that supported the new work environment. In our quest to do that, we found ourselves developing plans to develop plans, developing schedules to develop schedules, developing charters to develop project charters, hence the term: Meta-Planning. We were essentially developing the requirements of a new set of behaviours that we expected others to adopt. So, in our meta-planning efforts, we had many things to decide.

For example,

- Where were our trouble spots?
- What new behaviours or practices were needed?

- What was needed such as policies, templates, guidelines, to assist others in making process improvements?
- Who will develop the policies, template, and guidelines?
- When will they be developed?
- How will they look?

In short, after some false starts we learned that managing the software process improvement effort was a project in itself. It should be treated as a project. Organizing this 'meta-project' was difficult since there were few pre-existing, established and accepted practices, guidelines and procedures. The first step was to organize a management group to manage and oversee the project level work of all the Current Surveys. This group was an interdivisional group of managers from the programming division (ESMPD) and the planning division (EPCD), which consisted of the division chiefs, the Assistant Division Chiefs (ADCs), the StEPS Development Branch Manager and the StEPS Coordinator. This strategy built on the existing management structure of those responsible for the StEPS processing system. Everyone on the group had clear roles and responsibilities, and established authority for software planning and development. If the management group did nothing more than improve communications between the two divisions, it would prove valuable.

Our next step was to gather resources. Here we benefited greatly from an earlier Census Bureau initiative, the Software Engineering Process Improvement Group that was formed to develop a standard organizational software process to support software process improvement efforts across the Census Bureau. This staff had been working on CMM for quite some time. It was from this group that Ellen Soper, one of this paper's authors, joined the group as a process consultant to help on our software process improvement efforts. Her perspective was invaluable. One can learn much about tennis from reading a book about it, but consulting someone who has actually played a game is also necessary. We strongly recommend including a process consultant in any software improvement effort.

We also benefited from another ongoing program at the Census Bureau, which included extensive project management training and support. As mentioned above, CMM Level 2 is focused on instituting basic project management principles. We were able to build on this effort. After a while, we included an additional project coordinator who had gone through the project management training, and who had successfully coordinated other software related projects.

Still, we continued to try to co-opt and borrow as many resources as possible. Those who are actually managing and carrying out projects are very busy people. Learning and remembering to follow a new process is hard. Helping to develop processes, while still 'getting the work done" is too much to be expected. Managers must see "process improvement" as something that will make their work easier, not something that impedes the project. So we added a second process expert to work directly with one of Current Survey projects to help them apply the practices of good project management.

Just as in any project, the management group needed to define the work to be done for the software process improvement project by developing a project plan. And then, of course, the work was scheduled and available resources assigned. To address some of the problems mentioned previously, the management group agreed that improvements in the area of Software Project Planning and Tracking, Software Requirements Development and Management, and Software Configuration Management were necessary first steps.

We wish we could say that we followed our own processes and did everything "by the book." We didn't. One mistake we made was to combine our planning activities with the actual carrying out of

the activities. For example, we needed to develop a charter template, charter policy, and charter approval process. While we were developing these, we were also developing actual project charters and trying to get them improved. For several months, we worked on both aspects of our tasks as one project, with one schedule. This didn't work. We never had enough time to adequately discuss ongoing work problems and the process development work. In the confusion, a subtle but important aspect got lost. That is, the distinction between planning the process development (planning the plan) and following the processes developed. That is, we might discuss the schedule of when a requirements process draft should be ready. However, we never had time to discuss the substance of this process, at least in any detail.

The solution was, in hindsight, quite obvious, which was to separate the task of process development from the overall management of the current survey work. The same people were involved, but now we made the time to discuss the two efforts in separate meetings. One focused entirely on developing improved processes and the other focused on applying those processes.

Our guiding principle in the development and implementation of the processes was that they must be folded into our current environment and with existing resources. Obviously, a process "watered down" to the point of being ineffective would serve no purpose. On the other hand, a process that required too many resources would be rejected and viewed as interfering with "the real work." We wanted our initial efforts to be successful, but not burdensome.

It is important to look for allies, friends, and supporters. Everyone will say that they are in favour of the new process. After all, who could be against setting priorities, managing resources, and managing requirements? However, the fact is that the current culture and environment is considered to be working for some people. If you set explicit priorities and schedules, someone will consider these to be constraints. Someone who was getting his or her way now is not.

Our basic plan included developing core software processes, piloting the processes on a small number of projects to refine, providing training on the new processes, and implementing the processes on all current survey projects. In addition, we will establish minimum standards to set clear expectations. Our philosophy in developing the standards and processes is simple. They must be effective in achieving project goals, consistent with available resources, supported with training, and must be flexible to meet individual project's needs.

Progress Report

The management group spent the first year in developing an overall management plan, a process improvement plan, and has begun the work of documenting core software processes and templates. In addition to the planning, we have put in place some key management structures to address scope management, time management, and issue management. We have made some strides, but we have much more work to be accomplished in this area. However, we are learning, even in the early stages of our re-engineering effort.

A process where we have made great progress is in the area of software planning and tracking. We have drafted high-level process for Project Planning and Tracking. It was developed based on the recommendations of the CMM, as well as our own project experience and training. The process includes the development of a charter, the identification of a defined software life-cycle strategy, the documentation of the high-level system components, the development of written schedules, and sign-offs of key documents. We are now in the process of setting up pilot projects to implement the new process.

We have also developed a Project Charter Template to assist projects in following the project planning and tracking process. The intent of developing project charters is to document the project scope and to facilitate the agreement of the project scope. The process of developing project charters has brought issues and misunderstandings to the table for discussion and resolution. The earlier in the project these types of issues can be resolved, the better. It is easy for each of the user divisions to see each project from their own perspective and priorities, thus setting up expectations that cannot possibly be met. By requiring new projects to develop a project charter and obtain appropriate signatures on the charter, projects are communicating and resolving potential misunderstandings in ways never before accomplished.

Since our requirements flow in from a score of staffs in three different divisions, we are developing a Software Requirements Development and Management process. Requirements range from the trivial (e.g. indent a line, italicize a word) to the major (e.g. get the Quarterly Services Survey in the field by first quarter ’04). To be honest, we are struggling to define this process. The goal is to define and agree upon the requirements (baseline) and then control changes by assessing their impact and risk and, again, agreeing to make the change. This process should, of course, be done at the lowest practical level, which is easy to write down and hard to do. We have set up a sub-committee to define a workable process to address many of the issues related to poor requirements development and management. The Requirement Development and Management Process includes documenting high level requirements in a Users Requirement Document and documenting detail software requirements in a Software Requirements Document. This is a substantial change in how we have historically developed software; however, this is one area which all agree needs improvement.

We have drafted a Software Configuration Management process and a sub-committee is being formed to develop an appropriate plan deal with the software configuration including change control of the StEPS processing system.

A final, but quite major, piece of our work is developing an overall milestone schedule for the current surveys. This is a monumental effort because it requires identifying and defining the work of the current survey projects and summarizing that work in a meaningful way that can be managed. It requires establishing and agreeing upon clear priorities among the projects. We have instituted schedule review meetings for all chartered projects to give the management group insight into the status of projects so they can manage the resources and evaluate new project commitments.

7. Conclusions

CMM is a framework to develop improved processes to manage software development. CMM Level 2 reflects generally well-accepted practices of project management, coupled with software quality assurance. The Census Bureau has formally committed to CMM as evidenced by a Software Development and Maintenance Policy signed by Census IT managers. We are also committed to project management through the availability of a Project Management Certification Program that affords managers the opportunity to enrich staff through structured project management training. The Economic Directorate has begun implementing CMM-Level 2. We have generally found the model a useful framework that helps to organize our process improvement efforts. The processes that CMM suggests starting with do indeed correspond to the processes that we believe need improvement in our area. Although, we currently have no way to quantify our successes, we believe that CMM presents a useful framework.

However, even within that framework, there is much work to do. Initially, core staff needs to be trained on CMM and basic project management principles. Secondly, a steering group or management committee needs to be established to oversee the development and implementation of sound software practices. We recommend building upon existing management structures. The software process improvement initiative needs to be managed as a project. To be successful, this process must involve and be supported by senior staff. It must also include the involvement of the people who are expected to implement the processes; else they won't support the initiative. The notion of meta-planning is a necessary component of the re-engineering process.

As noted above, CMM does not mandate a particular process of software development, such as the "water fall" or "V," nor does it rely upon a particular set of software tools. It does mandate that these issues be explicitly addressed. Many organizations, including the Economic Directorate, had never seriously addressed these issues previously. Forcing this discussion is an important aspect of a software process improvement effort. However, the added time for this discussion must be considered.

Project management and process improvement can be made to seem as an unobjectionable and straightforward effort. In fact, nothing involving large groups of humans is simple or tidy. The current system, no matter how messy, is working for some people. Some people will have long standing working relations that allows them, effectively, to elevate their perceived needs to the head of the queue. An orderly process to determine priorities and control changes will mean that some people will no longer have their 'needs' met, they will be placed lower in the queue or simply be told 'no.' Do not expect these people to be supportive of the initiative. These are not insurmountable obstacles, but one should not be shocked or discouraged when they apply.

The second obstacle will be lack of resources, especially time. The Census Bureau has made a major commitment to CMM. Still, the task for the individual software engineers or project managers can seem daunting. They will still need to 'get the work done,' i.e., produce the survey. They will need to adopt new processes, which means taking the time to learn both the mechanics and the substance of the new process. This will be an added burden, with promised payoff in the future. And for those doing the meta-planning, they will have to develop the new processes, templates, training, etc. This investment in the future is, in our view, worthwhile. The alternative is the current chaotic system. However, it is an investment that takes resources.

Moving to CMM requires more than meta-planning. It requires a change in culture and a change in the business processes. Changing culture is the hardest part of any re-engineering effort. A necessary condition is for people to see and believe that the new framework of developing and managing software related projects is here to stay, and that the work they produce will be better as a result of it.

All paradigms for improving quality emphasize the needs for a strong management commitment. Fortunately, we are getting such strong management support, if only because the current situation is acknowledged to be unacceptable. The processes we are developing are increasingly being accepted and used. Some project leaders have adopted the entire package as the best way to ensure a successful project. Others have adopted only parts. With quality improvement, one can never say, "We have achieved success," only that things are improving. Thanks to hard work and senior management support, we are improving our process, although never as fast as we would like or dream. To summarize the moral of our story then, continuous improvement can be hard and hard to see. Don't get discouraged.

References

Adhern, Dennis M., Clouse, Aaron, Turner, Richard. CMM I Distilled. 2001. A Practical Introduction to Integrated Process Improvement. *New York: Addison-Wesley*.

Burwick, Diane M. 2001. How to Implement the CMM**.** BPS Publication, Ft. Knox, KY.

Caputo, K. 1998. CMM Implementation Guide: Choreographing Software Process Improvement. Reading, MA: Addison-Wesley.

Carnegie Mellon University, Software Engineering Institute 1995. Capability Maturity Model: Guidelines for Improving the Software Process, Addison-Wesley Logman, Inc.

Carnegie Mellon University, Software Engineering Institute, (No Date) Capability Maturity Model for Software, http://www.sei.cmu.edu/cmm.

Dymond, Kenneth M. 1995. A Guide to the CMM: Understanding the Capability Maturity Model for Software. Maryland: Process Transition International, Inc.

Humphrey, Watts S. 1989. Managing the Software Process, Addison-Wesley

Paulk, M., Weber, C., Curtis, B., and Chrissis, M. 1994. The Capability Maturity Model: Guidelines for Improving the Software Process. *New York: Addison-Wesley*.

About the Authors

Howard Hogan is Chief of the Economic Statistical Methods and Programming Division, Bureau of the Census, Washington, DC 20233. **Ellen Soper** is a Systems Analyst serving as a process improvement consultant to areas across the Bureau of the Census. This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Does the Questionnaire Implement the Specification? Who Knows?

Paddy Costigan & Steve Elder

Abstract

A recurrent theme since computer assisted interviewing became widespread has been the difficulty of representing the content of complex questionnaire instruments in a generally intelligible form. A key objective in development must be to ensure that the questionnaire instrument implements the researcher's intention faithfully. Automated systems for documenting the working instrument may assist this objective, but a gap may necessarily exist between an intelligible and unambiguous specification of a requirement and the detailed set of instructions that are ultimately necessary to implement that requirement in the software system used. An adequate specification that stands separately from the ultimate system instructions serves the purpose of providing a reference point against which the correctness of the instrument as implemented can be judged.

Keywords

Computer assisted interviewing (CAI); questionnaire instrument; question routing; dynamic text substitution; questionnaire documentation; quality assurance (QA); questionnaire specification.

1. Division of labour in the development of CAI instruments

Authors of and enthusiasts for particular software packages used for CAI like to hold out the promise that the researcher who makes or co-ordinates decisions about questionnaire scope and content can implement the required questionnaire unaided by technical experts. In this view the specification of the requirement and the detailed instructions required by the particular package to implement that requirement are often regarded as one and the same.

In practice it would be extremely difficult for a researcher to encompass all the skills needed to develop and manage a complex questionnaire instrument satisfactorily without being able to pass well-defined parts of the process over to a technical specialist.

Certainly much of a questionnaire specification can with advantage be expressed directly in terms of the package's own conventions and so be common currency between researcher and programmer or other technical expert. However, complexity can escalate rapidly in certain areas and to deal with such requirements satisfactorily and maintain a coherent overall instrument requires skills that are indistinguishable from programming skills, even if some are reluctant to accept that terminology in the survey context. Few researchers are likely to have the combination of interest, time and opportunity to develop and sustain all the necessary skills at the required level, alongside all the other calls on their time.

For these reasons, the National Centre for Social Research (NatCen) has evolved a form of collaboration between researcher and programmer that recognises the particular expertise of each, as well as acknowledging a large area of overlap in competence in the package used. NatCen has standardised on the Blaise system for all CAI questionnaires. Some features of the Blaise system assist this approach and some do not. The main issues that arise in development and that are dealt with here appear to be largely independent of the package used.

Before questionnaire development begins in detail at NatCen there is a meeting of all parties, the purpose of which is to establish the scope of the project, identify potentially difficult features, quantify the level of input that will be required by both researcher and programmer and agree a development timetable. Such discussion enables the overall structure of the instrument to be determined and the programmer to set up empty templates of the various modules to be defined.

Researchers are trained in the Blaise system and specify a high proportion of the questionnaire content directly in the form required by Blaise, delivering this to the programmer as a working instrument. However, some specifications of question routing as well as other control instructions are presented as formal specifications in words, to be implemented by the programmer. These specifications persist as comments alongside the appropriate question definitions and are amended subsequently if requirements change. Such specifications provide a route back to 'what was intended' and against which the correctness of the ultimate instructions to the program and its behaviour can be tested. Maintaining such specifications as comments is admittedly an additional task. It is clearly vital to ensure that such comments are always 'in step' with the program itself. However the task is a lesser one than trying to maintain the spec in a wholly separate parallel system (e.g. as a free-standing Word document). The latter may have the advantage of always providing the client with something readable, but ideally a readable version would be capable of being generated from the source program.

When all requirements have been implemented, both the working program and the 'source code' (the 'program' or set of instructions to the package with embedded comments) are passed back to the researcher for testing and potential amendment or refinement. Required changes are either made directly to the instructions in the source code, or further amendments are requested of the programmer by means of comments. This process may be repeated several times before and after piloting of the questionnaire. By means of agreed procedures for the management of files, the programmer retains overall ownership of the final code and can at any point determine precisely what changes have been made.

Once the questionnaire is signed off as fully tested and correct, its release to interviewers involves interaction with other systems that the programmer must initiate and to which the researcher's input is limited.

The above model for collaboration between researcher and programmer has evolved from and has replaced earlier practices. In the early days of CAI, researchers would draft questionnaires in whatever form they preferred, most commonly a word-processed document from which each text had to be lifted or keyed afresh. It was soon recognised that this led to substantial duplication of effort and a revised convention was encouraged in which question definitions were to be submitted in a 'Blaise-ready' format. This was an improvement, but take-up of the convention was patchy and it was also found that programmers still spent substantial amounts of time putting material into exactly the format required, which had often been neglected or misunderstood by the researcher. It was concluded that only full engagement of researchers with the Blaise software at specification time would ensure question specifications that were valid to the software.

In late 2001 a capacity crisis in programming forced the adoption of the model now described and a decision was taken by the management to make it mandatory with immediate effect. Rather than training researchers in the necessary software conventions in the abstract, Computing staff provided training at the outset of each project, after assessing requirements at an initial meeting. Resistance to the idea from researchers was limited, and largely evaporated as it was discovered that the new model was a different way of performing the same set of tasks rather than the addition of new tasks. It was also seen to have the advantage of leading more quickly to a first working version of the questionnaire, even if full routing control and other features were still to be added by the programmer. Programmers were sceptical and had to be persuaded that extra effort in training at the start of a project would bring benefits later, a promise that has largely been met.

This model of working became fully established during 2002 and attention has now shifted to codifying more detailed conventions in its use, particularly in the use of comments for communicating specifications and their later amendment.

Within the new model it is instructive to compare the different elements within the overall specification process, and to consider the extent to which the final instructions to the program can be regarded as an intelligible specification.

2. Definition of questions, answer ranges or categories

In the naming of questions and the specification of their content there is a largely a one-to-one correspondence between the elements in a specification as generally understood by researchers and clients, and the form of instructions required by a software package. This applies as much to quantities, dates and verbatim text responses as to the sets of coded answers that account for the greater part of most survey questionnaires.

```
HowMIG
"How seriously have you thought about applying for the
Minimum Income Guarantee? Have you thought about it…READ
OUT":
(VSer     "...very seriously",
QSer      "quite seriously",
NVSer   "or not very seriously?"
CSay     "SPONTANEOUS ONLY - (Can't say)")
```

The above example is close to being as intelligible as its equivalent on a paper questionnaire. An automatically reformatted presentation of the same instructions is likely to be acceptable to a client. Thus the ultimate form of instructions, or something generated automatically from it, can serve as a record of the requirement that is understood equally by all parties. Within Blaise the DELTA system, based on TADEQ[1], already provides this type of output and will conveniently suppress all the control characters, such as the brackets, inverted commas, @/'s etc. required by the program. Unfortunately it ignores the specification comments that may be necessary in more complex specifications, as they do not form part of the program instructions proper.

[1] The TADEQ project was funded by the European Union to develop a prototype of a neutral Tool for Analysing and Documenting Electronic Questionnaires, which would be made available to other interested parties including the developers and users of CAI systems. Elements of it have now been incorporated within the Blaise software, under the acronym of DELTA (Documentation using ELements of TAdeq).

There is one potential factor in question definitions that can make the text appear less intelligible when looked at as a record of what is to be asked. This is the use of 'textfills' – the dynamic substitution of text to fit the context set by a respondent's previous answers or other known characteristics. In Blaise the element that can vary must be given a unique name preceded by a set character ('^'). In the example below, the variable elements have also been highlighted in bold italics for clarity (although this is not a requirement of the software).

```
"Can I just check, have you ^or_partner been getting Minimum
Income Guarantee or Income Support at any time since we last
spoke to you in ^PrevDate ?":
```

Textfills are a powerful tool, but they mean that the question text is not fully meaningful until filled out with the values applying to a particular respondent in a particular set of circumstances. This effect can be extreme, as in the following example.

```
"SHOW CARD A.  ^Howabout  ^HomeHolX  ^DayBefor  ^PrevDay
^PrevDate ^DidYou ^HomeHolY ^ActCard":
```

Here the sense of the question cannot be discerned without further explanation. As with all textfills, there has to be an instruction or set of instructions to define how the text to be filled is determined. In Blaise such instructions are placed in a section (the 'Rules' section) that is quite separate from the question definition. Automated documentation schemes may present the two alongside each other, but how intelligible would many users find the following code - or any refined presentation of it - that an automated documentation systems is likely to manage?

```
IF (TripTyp = 1) THEN                          {i.e. Home-
based}
        IF J = 1 THEN       {i.e. first day}
           IF I = 1  THEN   {i.e. first `Trips' of the day}
              COMPUTE HowAbout := 'So, thinking about';
              COMPUTE HomeHolX := '';
              COMPUTE DayBefor := 'yesterday, (that is';
              COMPUTE DidYou := '), did you make any trips or
outings from your';
              COMPUTE HomeHolY := 'HOME';
              COMPUTE ActCard := 'for any of the activities
shown on this card?';
           ELSE              {i.e. Subsequent `trips' on same
day}
              COMPUTE HowAbout := 'Did you make any @Iother@I
trips or outings from';
              COMPUTE HomeHolX := 'HOME';
              COMPUTE DayBefor := 'on the same day, (that
is';
              COMPUTE DidYou := ')?';
              COMPUTE HomeHolY := '';
              COMPUTE ActCard := '';
           ENDIF
        ENDIF

        IF J <> 1 THEN {  i.e. days other than first day of
grid}
           IF I = 1  THEN   {i.e. first `Trips' of the day}
              COMPUTE HowAbout := 'And how about';
              COMPUTE HomeHolX := '';
              COMPUTE DayBefor := 'the day before, (that is';
           ELSE              {i.e. Subsequent `trips' on same
day}
```

```
              COMPUTE HowAbout := 'Did you make any @Iother@I
trips or outings from';
              COMPUTE HomeHolX := 'HOME';
              COMPUTE DayBefor := 'on the same day, (that
is';
           ENDIF
           COMPUTE DidYou := ')?';
           COMPUTE HomeHolY := '';
           COMPUTE ActCard := '';
        ENDIF
     ENDIF;
```

Note that this example doesn't cover the entire computation of the question text, omitting day-of-week and date (^PrevDay & ^PrevDate) which were computed separately depending on the day of interview, and were also used at other points in the questionnaire. Nor does it cover a second type of trip (trips from holiday bases, rather than from home), which required an equivalent set of computations.

At some point the researcher would have had to formulate a specification for what is required. In the example above, this is to some extent conveyed in 'comments' embedded in the source code, and shown here in {curly brackets}. Equally, this might well be naturally expressed in terms such as those that follow.

```
IF dealing with home-based trips (TripTyp =1), AND
     IF talking about 'yesterday', (J=1), AND
          IF talking about the first trip taken that day
(I=1), THEN make the text read as follows:
               "So, thinking about yesterday, that is (INSERT
Day & Date), did you make any trips or outings from your HOME
for any of the activities shown on this card?"
          OR
          IF talking about a subsequent trip on the same day,
THEN make the text read as follows:
               "Did you make any other trips or outings from
your HOME on the same day, that is (INSERT Day & Date)?"
     ELSE IF talking about a previous day (J<>1) AND
          IF talking about the first trip taken that day
(I=1), THEN make the text read as follows:
               "And how about the day before, (that is
(INSERT Day & Date))?" …….etc.
```

The above form of specification may appear to be redundant once the ultimate instructions are in place, yet this redundancy has value as a record of the requirement as conceived, and as a reference point against which the correctness of its implementation can be judged.

Despite the complications that textfills may bring, in general the specification of the content of individual questions presents fewer problems in communicating and documenting requirements and their implementation than the other control features now considered.

3. Specification of routing and other controls

In Blaise, instructions to control the sequence (or ‘route’) of questions asked of a respondent tend to account for most of the instructions in the ‘Rules’ section. Other instructions in the same section relate to the control of textfills, as described above, the activation of hard or soft checks on the correctness or consistency of answers, and other necessary computations.

Many questions may be unconditional and asked of all respondents. For many others the routing conditions may be simple and self-explanatory, e.g. those who answered 'Yes' or 'No' at the question immediately before. The use of meaningful short names rather than code numbers for answer categories, a feature of Blaise, helps to make the conditions convey their meaning.

```
HomeWarm
IF (HomeWarm = No) THEN
     WhyCold
ENDIF
```

Such examples are seductive, for they imply that it is a straightforward matter to formulate and assess routing controls of any complexity in terms of the program code ultimately used; however the complexity of expression can multiply very rapidly.

A further complication is that the conditions which govern whether a respondent is to be asked a particular question may form hierarchies. A local condition may be in addition to a more distant condition if a whole sequence already relates only to a subset of respondents, and in turn the whole section of the questionnaire may only apply to a larger subgroup of the sample. Hierarchies of conditions are a complication when attempting to communicate governing conditions intelligibly.

Consider the following definition, which determines whether a particular question - TotSav - is asked. It is not necessary to digest the detail, merely to register that the expression is long and complex. In practice this example has been truncated.

```
Holding[1] := ''; Holding[2]:= '';  Holding[3]:= ''
       FOR Count1:= 1 TO 7 DO
          IF Count1 IN [1..4] THEN
             IF (Count1 IN QAdInt1.Adult[1].Accounts) OR
                (Count1 IN QAdInt1.Adult[2].Accounts) THEN
               Holding[1]:=           Holding[1]           +
CAP(assname[Count1])  + ', '
             ENDIF
          ELSEIF Count1 = 5 THEN
……etc, for a further page, then

       IF (LEN(Holding[1]) > 2) OR (LEN(Holding[2]) > 2) OR
          (LEN(Holding[3]) > 2) THEN
          TotSav
       ENDIF
```

In this particular example the relationship to relevant question variables is increasingly difficult to follow, even if it can still be said to be explicit. What is lacking is any immediate and overall sense of the purpose and significance of the condition in the structure of the questionnaire. A comment of the following form provides both this sense and a reference against which the ultimate code can be evaluated.

```
There can be up to two adults in a Benefit Unit (BU), each of
whom has earlier been asked separately about their savings &
assets at the questions Accounts, Invests and OtInvA. Ask
TotSav only once per BU, provided that at least one of the
adults has any savings or assets.  Use textfills to build up
a list of all savings/assets held by either of the adults, to
be incorporated into the question wording.
```

In this example the solution to this dual requirement for routing and textfills is quite elegant. The complexity comes in constructing the textfills: these are empty to begin with, but are built up by complex references to earlier parts of the program, at a different level in the data hierarchy. The

routing condition for the question (consisting only of the final four lines of code) then simply checks if the textfill variables actually have anything in them.

Significant research inquiries often require additional dimensions of control in the interview process. Repeated questioning about a series of objects – journeys made, products bought, courses attended – is often necessary. Concurrent interviewing of several household members in one session may be necessary in order to limit the total elapsed time spent by interviewer and respondents. Such additional dimensions further complicate the expression of routing in the form that the system requires. The gap between an adequate form of specification that a researcher might naturally provide (and communicate with a client about), and the program code that is ultimately needed, becomes wider. In terms of the software system the best form of expression from the point of view of efficiency, parsimony and maintainability may look less and less immediately intelligible as it employs more and more programming features - temporary intermediate variables, loops, arrays, look-up tables and called procedures with passed parameters.

The following example of a specification cannot be faulted in clarity in the context of the particular survey from which it is drawn. It relates to the derivation of Family Units, potentially to be used for subsequent routing purposes, for instance where a section of the questionnaire needs to be repeated for each Family Unit in the household.

```
A family unit is defined as:

(1) a person living in the household on their own, (2) a
married/cohabiting couple on their own, (3) a
married/cohabiting couple and their never-married children,
provided that those children have no children of their own,
(4) a lone parent and their never-married children, provided
that those children have no children of their own, or (5) any
individual who cannot be assigned to any of the above
categories.

In general, families cannot span more than two generations,
that is, grandparents and grandchildren do not belong to the
same family unit.  An exception is when a grandparent is
acting in loco parentis and the child's parents are not in
the household. In such instances, provided the grandchildren
have no children of their own and have never married, then
they and their grandparent(s) will be treated as one family
unit. The same applies to great grandparent/child households.
The age of the child is irrelevant.

Children aged under 16 cannot be coded as cohabiting, but
they can be coded as parents.

Where two parents are present but are not coded as cohabiting
or married, then any children they may have will be coded to
the same family unit as the mother.  The father will form his
own family unit. The age of the parents is irrelevant :
offspring will always be grouped with the mother if she is
present.

Adopted and step-children belong to the same family unit as
their adoptive/step-parents.  However foster children should
be given separate family units from their foster parents.

Siblings (natural, step or adopted) do not belong to the same
family unit as each other, unless they have a parent or
```

```
 accepted guardian in the household.   Siblings who are
fostered together in the same household will be allocated to
family units separate from each other as well as separate
from their foster parents.

Same sex couples are coded in the same family unit.
```

Thus, for example, a household which included the HRP and his wife, their lone-parent daughter and her child and the HRP's mother would consist of three family units, the HRP and wife in one, their daughter and her child in another and the HRP's mother in the third.

Nevertheless it requires a large amount of program code - which we have truncated in the following example - to realise this requirement,. Whilst perhaps 95+% of circumstances could be encompassed relatively simply, 100% coverage including the rarest of cases necessitates code which runs to about three pages. Again, in this instance the reader does not need to follow the precise detail, but merely to register the complexity. Comments at strategic points within the program code help to indicate the function of each step.

```
{Compute Family Unit here:
NFam [1..10] family unit number for each person     NFamus -
incremented whenever new family unit needed}

     {Compute arrays NumPar and HasCh NumGP to indicate for
each person whether they have a child, parent, Grand parent
in the house. Used in next loop which computes the family
units.
Also need Mother/Father so can check for each child whether
Father spouse/cohab with Mother.}
     FOR Reloop1 := 1 TO DMHSize DO
        NumPar[Reloop1]  := 0
        HasCh[Reloop1]   := 0
        NumGP[Reloop1]   := 0
        Mother[Reloop1]  := 0
        Father[Reloop1]  := 0
……etc, for a further two pages, then
     {Now loop through the half-grid comparing each person
with all those before them}
     FOR Reloop1 := 1 TO DMHSize DO
        NFam[Reloop1] := 0
        {Person 1 is in family unit 1}
        IF (Reloop1 = 1) THEN
            NFamus := 1
            NFam[Reloop1] := NFamus
        ENDIF

        IF (Reloop1 > 1 ) THEN
          RelStop := Reloop1 - 1
          FOR Reloop2 := 1 TO RelStop DO
           {check relationships with previous people in grid.
If person passes any of the conditions below then they belong
in a family unit already used}

{Person 1 is allocated to Family unit 1. Then for each other
person check if they are in the same family unit as someone
who has already been allocated to a family unit. If not
create new family unit}
            IF (SpouseCo[Reloop1] = Reloop2)
               {INTERPRET AS: If Reloop2 is Reloop1's spouse
...}
```

```
              {Is reloop1 the child of someone in the
household and reloop1 never married and reloop1 has no
children and is not spouse/cohabiting with anyone}
               OR   (   (Mother[Reloop1]   =   Reloop2   OR
Father[Reloop1] = Reloop2)
                AND Single[Reloop1] = 1)
               OR……….etc
               {Siblings single and common parent in
household}
               OR  (   (QTHRels.QHRels[Reloop1].R[Reloop2]  =
Sib
                      OR   QTHRels.QHRels[Reloop1].R[Reloop2]
= StSib)
                AND …….etc
               {Related/unrelated, no children and common
grandparent in hhold but no parent}
                     {OR
(((QTHRels.QHRels[Reloop1].R[Reloop2] = Sib)
                      OR  (QTHRels.QHRels[Reloop1].R[Reloop2]
= StSib))}
               OR ……….etc
            THEN{Person belongs in an existing family unit}
                  NFam[Reloop1] := NFam[Reloop2]
            ENDIF
          ENDDO
        ENDIF {Reloop1 > 1 i.e. persons 2 to DMHSize}

        {Person not in previous family unit so create new
one}
        IF (NFam[Reloop1] = 0) THEN
          NFamus := NFamus + 1
          NFam[Reloop1] :=  NFamus
        ENDIF
     ENDDO
```

A researcher used to Blaise may well follow the logic of this and may make an important contribution to checking it. In fact, it is relatively easy to check whether the instruction has been *implemented* correctly, because it is possible to create scenarios covering all eventualities, the intended outcomes of which are known. However that is not the same as feeling comfortable in composing the code nor in using it as a basis for communicating what the program does to others. Nor would there be any advantage in attempting to spell out all the permutations in semi-algebraic English pseudo-code, as the risk of error in this process would be as great as in writing the code itself. It is partly a question of concepts that only become familiar if one works with them intensively, but also of the sheer number of steps that are necessary to realise the stated requirement. An alternative presentation of this same material, such as an automated documentation system would provide, is unlikely to change the picture significantly. It would require artificial intelligence at a level far beyond anything currently realised in survey applications to bridge the gap between the above two levels of description of the same underlying process.

Few researchers, let alone clients, will have desire, time or opportunity to develop the skills needed to assess fully the final form of instructions to the package, or any alternative presentation of them that a documentation system may provide. It is important to have a final version of key parts of the specification in a form that the client can understand and be willing to 'sign off'. Ensuring that the working program implements that specification then becomes a separate task of checking and testing.

4. Quality assurance of CAI instruments

Systematic testing of questionnaire instruments in their final form is often squeezed out of development timetables because the earlier stages overrun. Such testing as is done is compromised if it is used as a further opportunity to refine the specification. It then runs the risk of introducing new errors, as well as fulfilling its intended purpose of eliminating existing errors. True testing focuses on the question that forms the title to this paper: does the questionnaire implement the specification? NatCen have recognised the vital importance of true testing in their quality document. Nevertheless it is still difficult to get researchers to accept the nature of true testing in practice and the need to create and protect an adequate time slot for it in the development schedule, not least if faced with competing pressure from clients to accommodate last-minute amendments to the specification.

Even when there is a commitment to true testing it is difficult to disperse the task of systematic testing if there is no specification of the requirement that stands separately from the means of realising it. The output of automated documentation systems may be more friendly to the user than the ultimate program code, but it is nonetheless wholly determined by it and thereby is an inherent circularity. A separate specification is an additional layer that must independently be maintained and kept up-to-date, but provides testers with a benchmark against which to observe the instrument's behaviour. In the absence of such a specification, unless the tester is fully party to the objectives of the research and the specific intentions behind questions, there is no basis beyond common sense for commenting on the code or its behaviour. The program stands as the ultimate specification of what *actually* happens, but nothing can be said about whether this corresponds with what was intended.

Research teams often 'inherit' questionnaire instruments from other teams in the same organisation, or from other organisations when contracts for continuous or repeated surveys are re-let. Such questionnaires are rarely re-used completely unchanged, and it becomes necessary for *someone* to develop a full understanding of the program if it is to be adapted to meet further requests for changes. This process often throws up queries about what was the intention underlying the existing code.

In the dark corners of complex questionnaires it is not uncommon for a fresh eye to discover apparent oddities that may have gone unnoticed even during quite long periods of prior data collection. A routing instruction may appear strange in the context of the stated objectives of the survey, but who can say with confidence that it is wrong, in the absence of any independent specification of what was intended against which the program can be assessed? It is common for such instruments to be bereft of any comment or specification other than the detail of the program code itself, and the commissioning organisation - some years on and with changed personnel - may itself be none the wiser as to the detail of the original intent.

5. Conclusion

All of these arguments point strongly in the direction of needing a specification, independent of the program itself, which is intelligible to the non-programmer. This could serve two main purposes: (a) to develop the specification in the first place, and (b) to document the eventual program for posterity in terms intelligible to the lay reader. In developing a questionnaire, a suitable model might be that researchers and their clients would have responsibility for *agreeing the specification*, and the researcher and programmer would then have responsibility for *ensuring its correct implementation*.

Problems arise, however, if the specification has to be created and maintained entirely separately from the program code, in some format or currency acceptable to both researcher and client (such as a 'Word' document). This is a recipe both for duplication of effort and for potential error, since two source-files are involved.

A far better solution would be for the specification to be entered and maintained in a single source-file (that of the program software), with the facility (a) to insert 'plain English instructions' for specifying e.g. routing, textfills and computations, as well as the question text, and (b) to export these in an integrated fashion (preferably stripped of software-specific control characters etc), intelligible to the lay-reader. At the time of writing, however, we are not aware of an off-the-shelf package which fully meets these requirements. In Blaise, for example, the 'plain English instructions' can be entered in comments, but these are - by definition - ignored in all standard outputs, and there is no facility to associate 'plain English' descriptors with such items as routing conditions.

Whilst 'plain English instructions' will always have to be updated to reflect changes to the specification - a process which is unavoidably open to human error – this is nonetheless a much less arduous requirement than the maintenance of entirely separate source-files. However, ensuring that ‘plain English’ instructions and the ultimate program code remain in step with each other is still a real issue and must be addressed. There is no easy remedy for this problem, but at NatCen the emphasis is on codifying conventions for the communication and documentation of specification changes, and pressing the case for them to be universally followed. We feel that if these conventions are used as the basis for dialogue with the client and for what is finally signed off, that will provide further impetus for keeping specification comments updated, as well as providing further pairs of eyes to spot errors in them.

Automated documentation systems have now been under development for a number of years, following perhaps belated recognition by software developers that this constituted a genuine requirement. Some enthusiasts argue that such tools then have the potential to be developed in the reverse direction - i.e. as a specification tool which would drive an automatic code-generator. We are not arguing the case for this, recognising that they are likely to break down in the kinds of extreme scenarios we have discussed. However it is perhaps more surprising that - after so much development - it is still not possible to realise the more modest ambition outlined in this paper, namely the integration of 'plain English' specifications with the code needed to implement them, within the same documentation output.

About the Authors

Paddy Costigan is Deputy Director of the National Centre for Social Research (NatCen), where his responsibilities include the development of best practice and the streamlining of survey processes without loss of quality. He helped to pioneer NatCen's use of CAI in the early 1990s, and has had management responsibility for some of their largest CAI projects. He continues to head a quantitative research group handling a range of continuous and ad-hoc surveys. He can be contacted by e-mail at p.costigan@natcen.ac.uk.

Steve Elder was until earlier this year Computing Director at the National Centre for Social Research (NatCen). Now semi-retired, he continues to do some consultancy work for NatCen on systems for survey processing and on individual projects. As well as CAI specification his particular interests are in systems for managing fieldwork, data editing and metadata. He has been a member of the ASC Committee for many years and took over as the current Chair in 2002. He can be contacted by e-mail at s.elder@natcen.ac.uk.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

The Impact of Computer-Assisted Interviewing (CAI) on the Survey Process for Social Surveys in the Office for National Statistics (ONS[1])

Tony Manners

Abstract

CAI is widely acknowledged to be the most important technological development in the survey process in recent decades. ONS pioneered the use in the UK of CAI for large-scale complex household surveys. Its first CAI survey, the quarterly Labour Force Survey, started in production in 1990. Other existing and new surveys followed, and by 1995 ONS had moved over completely to CAI. Today, ONS remains at the forefront in the use of survey technology for interviewer-based data collection and processing.

The proposed paper will discuss the impact of introducing this major new technology in the UK's national statistical institute. One focus will examine the impact against the objectives of producing better quality data, faster and cheaper. Another focus will be on changes in the survey process, looking particularly at the strategic options and reasons for the particular choices made by ONS. The paper will make international comparisons with other national statistical institutes and agencies which carry out surveys for official statistics. It will look forward to the strategic issues of the coming decade.

Keywords

Blaise, CAI, CAPI, CATI, CASI, ONS

1. Introduction

This paper examines the impact of computer assisted interviewing (CAI) on the survey process for government surveys which are complex enough to need an interviewer to administer them. This classification is, of course, itself open to change from technological developments. Some commentators have looked forward to the elimination of the interviewer from the survey process , for example through increased use of technologies for passive observation[2]. However, the demand for interviewers for surveys for official statistics has been growing not receding. The main candidate for eliminating the need for interviewers, the web survey, is unlikely to fulfil this role in the foreseeable

[1] ONS was formed in 1996. In this paper, references to ONS in earlier periods are to the organisations from which it was formed.

[2] Baker, R. "The CASIC Future" in Computer Assisted Survey Information Collection, ed. M Couper et al, Wiley, New York, 1998

future, for reasons which this paper will examine later. In the main, it is the more complex social and socio-economic surveys which need interviewers, while business surveys tend to be sufficiently simple for self-completion to provide a reliable and more cost-effective method.

The paper examines the impact of technology historically rather than, for example, looking at each stage of the statistical value chain (SVC) in turn. In part, this is because a key impact is on the SVC as a whole, changing its elements and the meanings of the surviving elements in ways which are best explained as a process in time. A historical rather than functional approach can also bring out the crucial human interactions with technological possibilities which determine the nature of the impact on both survey technology itself and on the humans' ways of thinking about survey processes. New technology changes our thinking even more importantly than any direct impact it may have.

The core elements of CAI have been computer assisted telephone interviewing (CATI) and personal (i.e. face-to-face) interviewing (CAPI). The first developments in CAI took place in the 1970s and involved centralised CATI with mainframe and minicomputer systems[1]. CATI had a major impact on the survey process for relatively simple surveys, such as many in the fields of market research and political polling, where the speed with which a survey could be launched and the results obtained and published improved dramatically. However, most social surveys for official statistics did not use telephone interviewing of any kind, for methodological reasons such as its insufficient coverage of populations of interest, its inability to cope with lengthy interviews and concerns about lower response and other mode effects. Moreover, many official social surveys were too complex for cost-effective use of the CATI technology of the day.

National statistical institutes (NSIs) and survey agencies which carried out surveys for official statistics were much more interested in technology which would enhance the face-to-face interviewing methods that complex social surveys required. Couper and Nicholls have observed that NSIs led the CAPI revolution [2]. While there were other survey agencies conducting surveys which needed face-to-face interviewing, it was NSIs which had the considerable resources needed to undertake not only investigation but also implementation in production systems. Interest grew with the arrival of the personal computer in the 1980s but it took time for this technology to achieve sufficient mobility. One of the first investigators of the possibilities, Statistics Sweden, found that despite its serious efforts in the early 1980s, before commercial production of sufficiently powerful mobile computers, that it was not possible to have an adequate machine built to order[3]. It was only in about 1987 that computers achieved, in the MS-DOS 8086-chip forerunners of today's laptop computers, the combination of computing power, light weight and low cost that made CAPI feasible for production use for the more complex government social surveys. Nevertheless, by that year, Statistics Netherlands was ready with a pre-DOS system which it used to make its Labour Force Survey the world's first CAPI production

[1] Weeks, M " Computer Assisted Survey Information Collection: A Review of CASIC Methods and Their Implications for Survey Operations", Journal of Official Statistics, 8, 1992, pp.445-465

[2] Couper, M and Nicholls, W "The History and Development of Computer Assisted Survey Information Collection Methods" in Computer Assisted Survey Information Collection, ed. M Couper et al, Wiley, New York, 1998

[3] Blom, E et al "Computer Assisted Data Collection. Production Test in the Labour Force Surveys August 1989-January 1990", SCB, R&D Report, Statistics Sweden, 1990: 11E English Edition.

survey for official statistics, but it was already developing the DOS replacement, the first version of its now well-known Blaise software[1].

By 1987, ONS had already had the help of Statistics Netherlands in setting up its own quarterly Labour Force Survey on the latter organisation's first pre-DOS CAPI system so that "proof of concept" trials could be run in the UK. These trials, which were for both CAPI and quasi-CATI (i.e. without call scheduling), were a dramatic success in dispelling the then widespread expectations that there would be public resistance and that interviewers wouldn't be able to cope[2]. The public were, as one interviewer reported, "less than interested" in how the interview was captured. Interviewers, however, were enthusiastic. With hindsight, it is probably fortunate that the first trials used very light diskless machines; organisations which started "proof of concept" trials later, with more powerful but heavier machines running MS-DOS, encountered natural resistance from interviewers. ONS moved straight to the first relatively light MS-DOS machines, with the Toshiba T1000.

The trials established for ONS, once and for all, with buy-in from all functional areas of the interviewer survey process, that this was the way of the future. The buy-in was crucial because it meant that the process of developing a production system was not hampered by any part of the organisation feeling the need to preserve the old ways of doing things. For example, unlike the norm in contemporary trials in other countries in the late 1980s, ONS interviewers never carried paper and pencil back-up, as a matter of policy. Interviewers could focus on the new way of working, which in turn had to be (and was free to be) genuinely re-engineered to optimise the means for them to do so[3]. Questionnaires were redesigned for CAPI rather than being a direct import of paper and pencil design. This avoided the paradox of a more traditionally-based development approach which insisted on constraining CAPI design to something that could be directly measured against paper and pencil design - that this was setting up CAPI to fail. The organisational commitment meant that the development process met obstacles by taking more radical steps forward instead of taking a step back. In particular, ONS's field managers were fully supportive of the new technology from the very start and not, as in some other organisations at that period, a conservative force.

The trials also confirmed that CAPI development could and would be led by business area project managers, who did not have traditional information technology (IT) skills, because CAPI was based on a technology sufficiently mature that optimising its applications in terms of deliverables was more important than just getting it to function or optimising it in traditional IT terms[4]. Traditional IT skills were vital to the success of the development, particularly in such areas as building data transmission and case management systems, but this was always in a framework set by business areas, which also continued to be wholly responsible for questionnaire design and implementation. So the first, and in many ways the most important, impact of the new technology was to establish a mind-set that the survey process could be entirely re-thought not just tinkered with around the edges with some tasks

[1] Hofman, L and Keller, H "Design and Management of Computer Assisted Interviews in the Netherlands" in Proceedings of American Statistical Association, 1991 Joint Statistical Meetings, Alexandria VA, 1992

[2] Manners, T "The first laptop computer fieldwork project in SSD" in Survey Methodology Bulletin, 21, September 1987

[3] Blackshaw, N et al "Developing Computer Assisted Interviewing on the Labour Force Survey: A Field Branch Perspective" in Survey Methodology Bulletin, 27, July 1990

[4] Manners, T "Using Blaise in a survey organisation where the researchers write the Blaise datamodels" in Essays on Blaise, 1998, Statistics Norway, Lillehammer and www.blaiseusers.org/Ibucpdfs/1995-1998/manner98.pdf;

handed over to IT staff. In terms of deliverables, dramatic improvements in quality, delivery times and costs could be envisaged: these were to be the drivers.

The further impact of CAI in ONS can be seen in four subsequent phases to date. The major development phase focussed on getting the Labour Force Survey into production, first as the quarterly rotating panel component (achieved 1990) and then in the conversion of the whole LFS to this design (achieved 1992)[1]. The next phase involved the generalisation of CAI use to all ONS's other surveys, which was largely accomplished by 1995[2]. The funding regime of the 1990s meant that each move by a survey to CAI had to have its own business case, so the pressures were towards silo developments mitigated only by the most basic needs for interviewers to see approximately the same screen designs, in the same software, across the surveys they worked on; outputs, unconstrained even by such immediate operational needs, were developed ad hoc and reactively to the varying requirements of customers. As a result, the next phase focussed, whenever it could obtain resources, on developing effective ways of managing a large programme of surveys, particularly by standardisation and innovative approaches to providing the skills base for CAI survey design[3]. This phase, initiated in 1995 and continuing today, was supported by some of the major deliverables of the newly-formed Office for National Statistics: in particular, improved coherence in National Statistics through harmonisation and integration. These two themes demanded and provided means for better standardisation in the survey process, notably in the way CAI was used. A fourth, over-lapping stage, starting slightly later, has been to pick up on opportunities for quality improvements in the survey process which the move of CAI software into the Windows world has made possible from the late 1990s.

One of these opportunities is for web surveys, which some see as eventually removing the need for interviewer-based surveys and thus as a new technology as significant as CAI itself; others think it is clearly important for the field which already uses self-completion, such as business surveys and the Census of Population, but that it has a more modest role for interview surveys for the foreseeable future (for official statistics at least) as an additional mode to support CAPI and CATI in particular, constrained circumstances[4].

---

[1] Manners, T "Development and Implementation of Computer Assisted Interviewing for the British Labour Force Surveys" in Proceedings of American Statistical Association, 1991 Joint Statistical Meetings, Alexandria VA, 1992; "New Developments in Computer Assisted Survey Methodology for the British Labour Force Survey and Other OPCS Surveys" in Bureau of the Census, 1992 Annual Research Conference, 1992

[2] Manners, T and Diamond, A "Integrated Field and Office Editing in Blaise: OPCS's Experience of Complex Financial Surveys" in Bureau of the Census, 1994 Annual Research Conference, 1994. This shows that the percentage of all ONS interviews carried out by CAI rose from 50% in 1992 to 93% in 1994 and 100% (except for some small pre-pilot studies) in 1995.

[3] Pierzchala, M and Manners, T "Producing CAI Instruments for a Program of Surveys" in Computer Assisted Survey Information Collection, ed. M Couper et al, Wiley, New York, 1998

[4] Couper, M "The Promises and Perils of Web Surveys" and Flatley, J "The Internet as a Mode of Data Collection in Government Social Surveys: Issues and Investigations" both in ASC International Conference, The Challenge of the Internet, 2001.

2. Phase 1: re-engineering the survey process to provide results cheaper, faster and of better quality

In the economic and political climate of the 1980s, the top priority in the UK was to deliver the same products as before but cheaper. Although this does not sound inspiring, the effects may illustrate that what is important is to have strong drivers for change - which force people and organisations to take risks and innovate - almost regardless of their content. It was instructive at the time to discuss progress on delivering CAI with NSIs which saw it in more traditional terms as a means to improve quality first (and sometimes exclusively). They tended to find it difficult to justify the introduction of CAI because the quality benefits were hard to demonstrate when viewed in traditional manner, and since they were not rethinking the survey process specifically to deliver results faster and cheaper they actually found that CAI was more expensive and slower overall, so the cost-benefit case was poor.

ONS was able to deliver rapid progress because its main criterion - the bottom line cost - was very transparent and simple to measure, met immediate urgent needs of customers and, in the ideology of the times, was a conclusive argument inside the organisation against retaining traditional approaches. Of course, ONS also had to guarantee that the technology would work and would not damage quality or speed of results, but that is quite a different matter from finding quality improvements of sufficient magnitude to justify the risks and development costs. The net effect of these pressures was to force ONS (albeit willingly) to follow the full logic of the new technology, re-engineering traditional processes out of existence and avoiding introducing new ones, as tended to happen elsewhere, just because computers were now involved in the data collection.

The need to reduce risk meant a strong focus on building the simplest, most reliable systems, avoiding two common problems of the early days of CAI: (1) building rather than buying software; (2) over-specification and “cathedral-building” as many different interests sought to take advantage of the new technical possibilities (just one example: in choosing software we insisted on a simply and completely reliable ability to go back to earlier questions (“backing up”) and come forward on a new route if the respondent revised her views spontaneously or as a result of consistency checks; we did not therefore insist on audit trail facilities which some software offered to help sort out problems arising from unreliable backing-up). We chose to use an existing CAI package and to use it as we found it, rather than to try to force it to behave in the older, more familiar ways it was not designed for. The choice of the software was clearly critical. It needed to be designed on future-proof principles by people who understood survey methods as well as software engineering and to be simple and reliable to use. The Blaise software produced by Statistics Netherlands was the only package which, in our view, met these criteria; it has stood the test of time and is now the international standard for CAI for official statistics.[1]

In the first “LFS alone” phase of CAI, the major ostensible impact was on cost but this was based on underlying re-engineering of the survey process. As Table 1 shows, the business case promised to recover development costs within 2 years.

[1] See the website of the International Blaise Users' Group www.blaiseusers.org/

Table 1. Index of business case costs for PAPI and CAI options for quarterly LFS, 1989-1996							
Option	Financial years						
	89/90	90/91	91/92	92/93	93/94	94/95	95/96
Paper & pencil	100	100	100	100	100	100	100
CAI	102	123	83	83	94	89	83
Net extra for CAI	2	23					
Net savings for CAI			17	17	6	11	17

This was on target for delivery when the obvious success of the project led the customer to fund the re-development of the entire LFS from 1992 on the design of the CAI quarterly element (taking the annual number of achieved household interviews up to a quarter of a million from one hundred thousand).

Table 2 shows, for a year of medium savings where the cost savings in the new system came from: elimination of processes, particularly automation of clerical processes (such as keying, office coding and field management tasks like booking-in paper questionnaires). The obvious savings in printing and postage were less important than elimination of office coding and editing and of management and training for fieldwork elements which were no longer needed.

Table 2. Index of components of CAI business case costs for a year with some hardware replacement	
Cost of paper and pencil	100
Gross extra costs for CAI (computers/training/consumables)	10
Gross savings from CAI	21
Clerical coding staff	7
Clerical field management staff	6
Clerical keying staff	5
Other	3

Some radical re-thinking of the survey process (radical enough that even today, no other organisation has adopted it) which preceded CAI helped to realise fully the advantages of the new technology: the choice to use interviewers to code occupation and industry texts. The quality and timetable advantages of pushing editing upstream in the survey process to the interview were widely acknowledged, but timetable advantages tended to be reduced significantly where the residual requirement for coding verbatim answers left a coding stage. At a minimum, most official surveys needed occupation and industry coding. The accumulation of cases for coding by a small team is a potential bottleneck in the survey process and runs against the grain of the new technological possibilities. The wish to remove this stage had led ONS in the early 1980s to investigate the methodological issues in having verbatim answers coded to standard classifications by interviewers: the conclusion was that there were quality advantages, essentially trading increased variance for reduced bias, despite the conventional wisdom

that using "expert" coders was the only conceivable approach[1]. ONS implemented interviewer coding of classifications like occupation and industry. This decision went with the grain of the new technology and looked forward to the use of computer assisted coding (implemented shortly after the introduction of CAI when the Blaise tool became good enough at matching) and to negotiating the code with the respondent in the interview (as in "do you recognise yourself in this category that I'm coding you to") which is the next logical step in the search for more accurate data.

3. Phase 2: extension to all surveys, including the most complex

The success of the LFS implementation and its re-engineering of the survey process led from 1992 to the policy of using CAI for all new surveys, including the first British Crime Survey, and the conversion of the continuous surveys at the first opportunity. The complexity of the LFS, with its panel design requiring data from the previous wave to be available at each question in the current interview and the requirement to move interviews in both direction between CAPI and CATI, should not be underestimated. Nevertheless, the sheer size and the larger number of hierarchical levels in some of the surveys presented a capacity problem for the CAI of the day. The National Travel Survey had 7 levels and the General Household Survey up to 5 (depending on topics included).

The Family Expenditure Survey only had 3 levels but still had 6,000 unique questions and 30,000 questions overall, plus thousands of edits and derived variables, for the interview alone and also had two-week expenditure diaries. It also demanded concurrent interviewing (ability to interview all household members at the same time, each with their individual paths through the questionnaire) to ensure that shared expenditure and income were neither omitted nor double counted. On the principles of uncompromisingly eliminating processes, on the lines established by the LFS, the FES CAI redevelopment put the entire mainframe batch edit into the Blaise questionnaire with the net result of more than halving the timetable from fieldwork to results and saving one million pounds a year.

The reduction in the timetable would have been greater but for this survey there proved to be a non-technological limit to how much could be done in the interview, so a small amount of office editing had to be retained, together with a capacity to capture and code data from the paper diaries, with a disproportionate effect on the speed of processing. The complexity of a small number of important edits for the FES meant that a limit on resolving inconsistencies in the interview was reached when (1) the respondent did not have the knowledge required - for example she might know the overall amount received for multiple welfare benefits but not their precise make-up, so expert knowledge of the benefit system was needed[2]; (2) even if the respondent could possibly work out an answer, the amount of time it would take to resolve a dozen or so items involved in the check would be disruptive to the interview as a whole. Nevertheless, the whole processing system was carried out in Blaise, covering interview, intelligent data entry and coding of diaries and office editing[3]. The solutions to the capacity

[1] Dodd, T "An assessment of the efficiency of the coding of occupation and industry by interviewers", New Methodology Series No.14, 1985

[2] Further automation could, of course be envisaged. Attempts have been made, outside ONS, to use neural networks for editing and imputation of similar kinds of survey data but have not to date provided useable results.

[3] Manners, T and King, J "A Controlled Trial of CAPI and PAPI modes Data Collection and Editing: Estimates from the UK Family Expenditure Survey in February 1993" in Proceedings of the International Conference on Survey Measurement and Process Quality, American Statistical Association, Alexandria VA, 1995

problems mentioned above involved some ingenious case management design by ONS's IT specialists to provide a framework for the Blaise system.

4. Phase 3: standardisation and harmonisation for an integrated programme of surveys

As it became clear that all surveys, despite their range of different commissioning conditions and sponsors, and their different business cases, would use CAI, a framework for efficient management of a programme of surveys was created. ONS interviewers may work on several surveys concurrently, so the most urgent need was for a case management system which, at the laptop end, was flexible enough to ensure that she could manage her workload in the most efficient way and provide the right survey for whichever address she needed to visit next; to allow her to do whatever coding and administration was needed after the interview; and to return completed interviews quickly and in an orderly and accountable fashion to the office data handling systems.

At ONS, James Gray and his IT colleagues designed a system (known as CaseBook on the interviewers' laptops, and Scatter and Gather in the data handling and transmission elements) based on the idea of managing objects which could contain any kind of information and/or methods; in the main these would each contain all the information and methods needed for data collection at an individual sampling unit[1]. Each object could have unique contents. Among the many benefits of this idea were (1) any number of surveys, with any characteristics, could be managed efficiently; (2) the flexibility to allow any individual case to pass through processing at its own optimum speed, avoiding the bottlenecks common in paper systems and early CAI case management systems which tended to accumulate cases at each stage before they moved forward together to the next stage (where once again they queued for the next stage). One result of the new system was to take full advantage of the removal of the batch editing stage (already realised in the LFS development) by continuous processing (such as creation of derived variables back in the office) so that at the end of a field period only relatively few cases, from the last few days, remained for completion. This allowed the LFS to deliver population-weighted quarterly results for 60,000 households and more than 130,000 adults, with some 400 derived variables, to the ONS analysts in 4 days from the end of fieldwork instead of 7 weeks. This "object-based" approach has since been adopted by several other NSIs.

Standardisation of questionnaire design across surveys was equally important, particularly at the sharp end where interviewers needed to see screens which gave them the tools for interviewing in a reliably expected way, through standard screen layouts and conventions: for example, the name of the household member currently being interviewed should standardly be displayed on every screen and always in a standard place on it[2]. Rapid and accurate development, testing and maintenance of Blaise instruments required strong authoring standards and, wherever possible, modules of standard, well-tested code which could be plugged into new surveys (e.g. for the standard classificatory data), as well as templates embodying the best practice and solving design problems so that there was no need to wrestle with them anew on each project.

[1] J Gray "An Object-Based Approach to the Handling of Survey Data" in Essays on Blaise 1995, ed V Kuusela, Statistics Finland, 1995 and www.blaiseusers.org/Ibucpdfs/1995-1998/gray95.pdf

[2] Bushnell, D "From DOS to Windows: Usability issues for Interviewers" in Proceedings of the Third ASC International Conference, Chesham, 1999

Implementation of these standards and templates allowed ONS to continue to use the researchers who led the projects (defining the research problems, designing the means to investigate and measure them, overseeing the data collection and analysing and reporting on the data) to exercise their methodological skills in questionnaire design (as they had in paper-based systems) in the new electronic medium[1]. The ONS view was that designing on paper for programmers to implement electronically was unnecessary duplication and, above all, did not meet the specific design skills (neither design for paper nor computer programming as such) needed for the new technology. The Blaise software is simple and powerful enough for researchers to use unaided, as they do statistical analysis packages (in fact using much the same logic). The issue of standards did not arise in the first few CAI projects, but when some 50 researchers needed to be able to design Blaise instruments at the relevant stage in their projects it became essential. Researchers move between projects and work in teams so they need to be able to understand each other's code, so no idiosyncrasies or individual styles are allowed; code is written to be maintainable rather than computer-efficient since the former is vital and the latter an irrelevance in this context. Standardisation in ONS fostered and was fostered by the National Statistics project , beginning in this period, to harmonise concepts and questions for government social surveys; and by initiatives such as the prototyping of an integrated household survey which brought four major continuous surveys into a single CAI instrument and sample design.

As all surveys used CAI, interviewer training changed to concentrate on this mode. From 1995, ONS ceased to train interviewers in how to administer paper questionnaires with their now obsolete skills requirements such as following skip patterns accurately. Instead, more time could be spent on how to make best use of new electronic facilities such as context-sensitive help, available on a hot key, on the substantive issues in the topic under investigation (as opposed to the traditional method of memorising detailed subject information imparted in briefings and interviewer instructions on paper). Another area of major quality improvements made possible by the new technology was in the centralised CATI unit, where monitoring software allowed supervisors to target their oversight of interviewers' performance, either for particular interviewers or for particular sections of an interview across interviewers. Questionnaire designers spend much time designing standard measuring instruments but much still remains in the hands of interviewers: centralised CATI provides a higher level of control and feedback. A key development for the next few years is to implement such quality monitoring in real time for face-to-face interviewing: it has long been an aim - I can remember Bill Nicholas at the US Census Bureau talking about the vision in the early 1990s - but is now becoming feasible from a cost as well as technical viewpoint.

5. Phase 4: further quality improvements

The move of heavyweight CAI software into Windows in the second half of the 1990s was a longer and more difficult passage than the users anticipated. It turned out that some packages did not have this future designed into them. The Blaise package did - its final DOS version, Blaise III, had faced many of the difficulties early because it was a complete re-write with Windows in mind. As a result, the transition for Blaise was straightforward and it has become the international standard software for CAI for official statistics for social surveys; and it is expanding rapidly in the business survey field.

[1] Manners, T "Using Blaise in a survey organisation where the researchers write the Blaise datamodels" in Essays on Blaise, 1998, Statistics Norway, Lillehammer and www.blaiseusers.org/Ibucpdfs/1995-1998/manner98.pdf;

The Open Blaise Architecture, with tools like Blaise Components, is a key development which enables users to build their own extensions to functionality - such as a specialised interface for computer assisted self-completion or an alternative web survey tool; an early ONS application was a tool for analysing changes between questionnaires (now provided for within Blaise). A recent (2003) workshop hosted by the US Census Bureau and the Bureau of Labor Statistics asked a range of experts if there was any alternative to Blaise in the foreseeable future for surveys for official statistics and none suggested that there was.

From ONS's viewpoint, the current situation has many advantages. The software is developed by a sister NSI, Statistics Netherlands, with whom we have a long and very fruitful partnership on this and many other matters. There is a huge (for this niche market) international user base of NSIs and survey agencies which contract for official statistics, with thriving formal and informal mechanisms for exchange of views and information. In particular, there are formal mechanisms for user involvement in formulating strategy which have helped guarantee that Blaise's strong record on innovation has continued in the absence of effective competition since the late 1990s.

The impact of CAI's move into Windows has been to consolidate the changed survey processes of earlier phases. In ONS, as in its international peers, the new developments have been more to do with tools for improving the quality of survey design which the rich Windows environment makes possible, and with tools for further automating the interfaces between stages of the process (such as the Blaise tools for linking with data repositories and for web-based metadata documentation), than with dramatic changes to the process itself.

ONS has concentrated its new development efforts in two areas. Firstly, with many others, it is investigating the potential for various forms of audio-based self-completion (audio-CASI) in which the respondent listens to questions on headphones (for confidentiality) and keys the answers into a linked laptop. The routing of the questions is determined by the CAI instrument in the normal way. There is good evidence now of improved accuracy from audio-CASI in surveys where the questions are sensitive and confidentiality (even from an interviewer) at a premium; and in surveys where self-completion is desirable but literacy is a problem in the population concerned. ONS first used audio-CASI in 2001 in a study of looked-after children[1]. Secondly, ONS has been using the Blaise "intelligent" audit trail tool, with the help of a further tool from Statistics Canada, to investigate and analyse the micro-processes in interviews and self-completion elements to feed back into improved survey design[2].

However, it may be that a larger impact than these specific developments will come from the continuing consolidation of re-engineered processes. The vastly increased computing power from hardware and software improvements of recent years means that the logic of carrying out as much processing as possible as far upstream as possible can be followed through without technological limits. It is now normal for the "interview" instrument to contain much more metadata , and to do much more processing in the background, than is needed in the interview. The instrument ("datamodel" in Blaise terms) is a convenient single place to define (or reference external sources for) any metadata that the survey may need in the interview or later in the analysis. In particular, we have

[1] Gatward, R "Audio-CASI with Challenging Respondents" in Essays on Blaise, 2001, Westat Inc, Rockville MD, 2001 and www.blaiseusers.org/ibucpdfs/2001/Gatward--IBUC_Paper--Final.pdf

[2] Bumpstead, R "A Practical Application of Audit Trails" in Essays on Blaise, 2001, Westat Inc, Rockville MD, 2001 and www.blaiseusers.org/ibucpdfs/2001/Bumpstead--IBUC_paper.pdf

always derived variables which are needed for routing or checks during the interview but now we are moving towards deriving all standard and planned microdata-based derived variables in the CAI instrument where they will be self-documenting, available in well-tested standard code modules and available for analysis faster than if created at a later stage; it also happens that Blaise handles complex data structures better than standard analysis packages so it is a more efficient tool. The recent completion of an effective tool for displaying the metadata from complex data models (TADEQ) is a further incentive to make maximum use of the Blaise datamodel capabilities.

Two major technological developments - the web and mobile phones - need to be mentioned though they have yet to make much positive impact on the social survey process for official statistics, through CAI or otherwise. The obstacles to obtaining samples, of any but some special populations, which are adequately representative for the purposes of official statistics means that for the foreseeable future ONS sees use of the web for data collection in household surveys as restricted to a minor part of a mixed mode strategy. Even then there is much methodological work needed to establish its best design and uses, as has been done for other modes. That work is in its infancy, in that evidence and practice has not yet arbitrated between diametrically opposed views among leading methodologists. Mobile phones, with or without internet access, currently present a threat as much as an opportunity in the CATI field since they further undermine the viability of random digit dialling by adding coverage and selection probability problems to those of poor response rates. They can help in reducing non-contact rates on conventional surveys.

6. Conclusion

Couper and Nicholls[1] have argued that CAI has mainly been used to do the traditional processes more cost-effectively and faster, rather than to improve quality by dramatically different methods. The impact of CAI at ONS has indeed been to improve surveys dramatically in the readily measurable fields of cost and speed of delivery; there have been more incremental but important improvements in quality[2], which the major modernisation programme of the next few years will take forward faster.

About the Author

Tony Manners works for the Office for National Statistics and can be contacted at tony.manners@ons.gsi.gov.uk.

[1] M P Couper & W L Nicholls II "The History and Development of Computer Assisted Survey Collection Methods" in Computer Assisted Survey Information Collection, ed. Couper et al, Wiley, New York, 1998

[2] W L Nicholls, R P Baker & J Martin "The Effect of New Data Collection Technologies on Survey Data Quality" in Survey Measurement and Process Quality, ed. Lyberg et al, Wiley New York, 1997

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

The Metadata Dynamic. Ensuring Data has a Long and Healthy Life

Simon Musgrave

Abstract

The demand for intelligence from policy makers, analysts, researchers and the general public is growing all the time and data is a fundamental component. These data have to take a long and sometime tortuous path from conceptual modelling and data collection processes through actual raw numbers to information and intelligence in order to be of use. Rather than seeing data as a number of static data files on the one hand or as part of a complex database on the other, we need systems that manage metadata such that they allow data to flourish by moving freely through their different manifestations, whilst knowing where they have come from and where they might be able to go. Given the growth of web based data and content management systems, the final part of the presentation will examine how some users are building systems consistent with modern content management and advanced data browsing tools. It is essential that the perspective of both the data user and the data intermediary is fully accommodated so that the data web can be developed with consistent and clear meaning all the way from data conceptualisation to end use.

Keywords

Statistical metadata, secondary analysis, content management, data dissemination

1. Introduction

Data is a vital ingredient for the creation of information architectures and knowledge products. However to ensure consistent understanding and proper use of data, it is necessary to have metadata systems that are consistent, comprehensive and flexible. They should aspire to meet the needs of the data producer (whether a survey organisation or administrative authority), the data publisher (who may be bringing together disparate sources or creating summary tables) and the end user (who may have little connection with the original source).

Every user in data life-cycle will have a different, and potentially contradictory, view of what metadata is and what it is for. For some users they are all about controlling and managing the production process, ensuring efficiency and consistency of procedures. For others they are about understanding precise definitions and meaning so that data compilations (such as international data series) can be produced. For others they are about resource discovery so that the data can fit into wider resource catalogues. Other users will be interested in the context and meaning of the numbers so that they can be re-analysed and enhanced.

These growing uses and users do demand a wider understanding of metadata issues and so we will examine the different types of metadata and their use with the aim of understanding inter-operability issues. The perspective will be driven by the needs of the data publisher and end user but will cover some issues raised by different aspects of metadata, including conceptual, statistical, data management and administrative (see Froeschl et al 2003).

2. Definition

Amongst the many definitions of metadata, in order to focus our thinking on the statistical domain, a good definition is that of Westlake (2003) 'Statistical metadata is anything that you need to know to make proper and correct use of the real data, in terms of, capturing, reading, processing, interpreting, analysing and presenting the information.

Thus, metadata includes (but is not limited to) population definitions, sample designs; file descriptions and database schemas; codebooks and classification structures; processing details, checks, transformation, weighting; fieldwork reports and notes; conceptual motivations; and table designs and layouts. In other words, anything that might influence or control the way in which the core information is used by people or software.'

Thus we see that metadata covers an extensive spectrum of information and has to handle a broad variety of meaning. As statistics are often about life in the broadest sense, they have to cover meaning in both the domain of analysis as well as the statistical terminology itself. This paper concentrates unashamedly on the needs of the data user and focuses on later stages in the data life cycle. Consequently the metadata that are of most interest are those related to 4 through 6 in the list above but include information related to the full life-cycle, hence 7 through 13.

3. Purpose of metadata

Users of metadata

Metadata in the era of the web has two main audiences, the human and the machine. For the human user, there is a multiplicity of roles. These range from the professional who conceptualises a data model all the way through to the public reader of a data table in a newspaper. Their needs are very different and even amongst professional users, their understanding of metadata is driven both by their role and by the particular professional background. For example the computer scientist is likely to view metadata very differently from the social scientist (different conceptual background) and the medical statistician is different from the political pundit (different domain) even if all of these are carrying out a similar function.

What is particularly noticeable to the author is the rapidly growing number of intermediaries who are taking data from a variety of sources and want to present them in their own information and presentational context. Consequently it is not sufficient to have some easy to use central data dissemination systems, but rather the ability to extract the central information, ideally automatically, in such a way as to merge it seamlessly with other, maybe local, sources. Indeed as I will discuss later, it is not just linking to other data that is important but to the presentation of a wide variety of other information and intelligence resources.

Also highly relevant to metadata systems is the fact that an increasing number of users are viewing statistical data, most of whom have little or no background in statistics even at an elementary level. Hence the demand for greater clarity probably exceeds the capacity of the data publisher to either comprehend or service this demand directly themselves.

Semantic web

For both the human and the machine user we are entering the world of the semantic web. According to Hendler et al (2001) 'The semantic web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work together'. The web has made the publishing of textual and graphical information easier and cheaper than ever. Vast amounts of information have been made available worldwide at a press of a button, at virtually no cost and in a highly integrated way. By applying the principles of the Semantic Web statistical metadata, data and knowledge can be made just as easy to publish, locate and access. Assini (2002) describes how the semantic web makes it possible to create a new class of distributed applications. Statistical information is made available as objects described in such a way that remote applications know what they can do with these remote objects. In this way the use of urls enable users to return to statistical and metadata operations in real time. As a result a knowledge (or simply data) web can be created that integrates very heterogeneous material in one homogeneous presentation.

What is significant about the web technologies is that they provide much more flexibility than a traditional database. Whilst a database may be more efficient in terms of retrieval and modelling of a single data environment, it does not normally have a focus on the explicability of its content. A good metadata system should have a good model behind it that enables an understanding of the content to be placed in the right context.

Data Production

The key point underlying this paper is that data have a life well beyond the original production cycle. Froeschl et al (2003) define a linear data sequence for the production of data (not dissimilar to the points in Westlake's definition) as:

- Definition and design
- Production
- Process

The focus in this paper is on stage 4 and I shall argue that this should be seen as a much more recursive process, which has the potential to impact earlier stages in the process. To look at this another way, data are no longer isolated static objects. The fact that they are viewed this way is, to an extent, a function of the history of computing as numbers were the only part of the knowledge production cycle that used computers in the 60s and 70s. Words became a major part of IT in the 80s and 90s, but it is only with the extraordinary flexibility of the web that we are able to build completely open-ended information systems that allow data to flourish in their proper context, i.e. as an integrated and inter-operable part of the production of knowledge, rather than a discrete resource standing on their own. So today we can see that a data source at the end of the production cycle can do one of several things:

- They can merge with other data sources and become part of a bigger whole (for example in Eurostat data series)
- They can gradually be enhanced by attracting new derivations, additions and links.

What is needed is a metadata system that takes a downstream view of data usage and allows the end user to add and append data and other types of information to the original sources. Of course with a data web, it is possible for appendages to be located wherever necessary and also for one appendage to be used by numerous other data objects (for example an effective thesaurus or classification system).

Figure 1 (from Ryssevik, 2000) puts this data life cycle concept in visual terms. It shows how we need metadata systems to bridge the gap between the production process and all types of intermediate and end users and facilitate communication both down and back up the production process. Without an adequate focus on the secondary users, we risk under-exploiting the value of the resource.

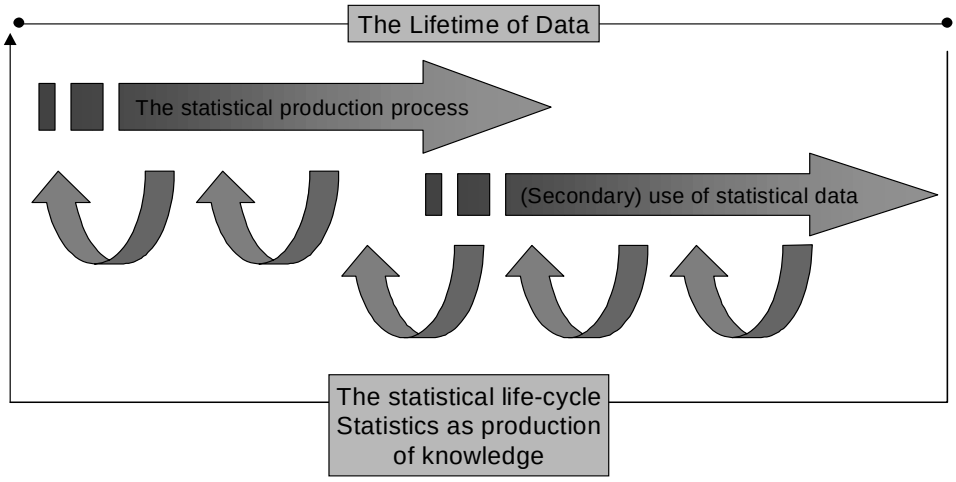


Figure 1. The data life-cycle

The production process corresponds to parts 1 to 6 in the earlier definition of metadata, i.e. everything from capturing to presenting the information. On the other hand the secondary use of data focuses on parts 4 to 6, interpreting to presenting the information. If we ignore this part of the life-cycle, not only do we lose much of the productive potential of the data, but we also lose the opportunity to improve the earlier parts of the production process by ignoring quality feedback.

4. Potential of Metadata

With an appreciation of the requirements of the end user and an understanding of the data life cycle, the fundamental requirements for the metadata become clearer. They are to facilitate the integration of the information garnered from the statistical system into a useable framework consistent with the conceptual domain. An example is the use for health survey data within a planning and decision making environment that drives the prioritisation of the creation of new physical resources. Another example is the integration of simple economic indicators into a dynamic presentation of the relative strengths of a European region.

Linking to other resources

In the growth of the vision that inspired the development of Nesstar software (Ryssevik and Musgrave, 2001), we have been driven by the concept of the data workbench The analogy is to the carpenter's workbench at which he brings together all the building material (raw or semi-processed), instructions and designs, tools and finishing materials. In a similar way the policy analyst, researcher or other data user wants to bring together disparate data sources of different kinds, different texts, for example on

methodology or the subject domain, a wide variety of data management, analytical and presentational tools and the ability to interact with a large number of external colleagues.

Content management

Major developments underway within government (in the UK and elsewhere) are centred on the development of content management systems. These are relatively high level systems aimed at separating the content from the presentation of, typically, textual information. There is a focus on metadata descriptions, as defined by the e-GIF (Government Interoperability Framework). The metadata descriptions (in the e-GMS schema) can be considered as an extension to the Dublin Core. Whilst it is a very useful addition for the location and high level classification of resources, it does not help very much with the types of statistical metadata needed for the correct understanding and presentation of statistical information.

What is important is the way in which it can provide part of the bridge to other types of material. This material includes reports and policy papers (especially at the public sector) and the percentage of these reports that include, at one level or another, some statistical material is open to conjecture but is likely to be very substantial, maybe 50%. However the presence of an e-GMS record does not allow us to link back to the underlying sources, or forward to other manifestations of the same information. To do that we need to develop more integrated systems, potentially exploiting some of the semantic web technologies described earlier.

Other Resources

It is worth reviewing some of the other types of information that could usefully be linked to statistical data and so note the types of metadata that are needed to provide meaningful and effective links. On the one hand there are resources that will help in the display and understanding of information, such as maps. On the other hand there are complementary resources such as

- Other types of data, (e.g. indicators, aggregates, individual level data and classifications)
- Expertise
- Training

In order to access these resources there has to be some agreement in the structure, syntax and semantics of the metadata and so any progress on these issues is helpful. It is generally recognised that XML is the most useful syntax but we are still struggling on the structure and semantics as discussed later.

A particular type of information worth reviewing is the linking of indicators or targets to underlying data. Often for the presentation of information, particularly on the web, the key facts might be placed in a framework or architecture. A good example is the Local Futures architecture. These are essentially displays of key performance indicators within an easily understandable context. Indicators can, and probably should, be linked to underlying data systems that enable the end user to interrogate and understand their data context. In addition some users would like to include the targets in data tables so that progress towards them can be viewed easily. By taking this approach we see that there are multiple entry points to data, via traditional catalogues as before, but also browselists and hyperlinked information space that provides added value route-maps to data.

5. Metadata Models

The focus of the paper is on the life-cycle of data and their potential to flourish or die. A fundamental aspect of this data environment is the underlying model in which data are placed. There are many data and statistical models, for example ISO 11179 and the Object Management Group's Common Warehouse Model (CWM). For the uninitiated they can be rather intractable, but essentially the models are seeking, for different operational or intellectual reasons, to understand the concepts, processes and inter-relationships, in other words to define behaviour of elements and not just the structure. Indeed one of the impacts of the web is to break down barriers between information systems. Where they are completely independent, the underlying model matters less than whether the system does the job required. As the systems develop and seek to be understood (one of the requirements for the semantic web) then the underlying model becomes increasingly important.

As well as the underlying model being understood, the terminology used must be agreed, understood and consistent. The use of common classifications is important for consistency. The classifications, for example the International Classification of Disease and Standard Industrial Classifications, are used in a system. This is then more efficient, in terms of storage, and enables the system designers to concentrate on the view of those classifications, rather than trying to develop them.

Another important point about classifications is that simpler taxonomies (or controlled vocabularies) can be constructed from them to allow end users to navigate them for resource discovery, and data publisher to use them for classifying resources. A good example is the Public Health Information Tagging System (phits). This has an additional, and not to be overlooked, pedagogical role as it enables users (particularly the less experienced) to find their way around the domain more easily.

Unified Metadata Architecture for Statistics (UMAS)

Froeschl et al (2003) have proposed a Unified Metadata Architecture for Statistics (UMAS). The goal is to ‘Define a framework to understand communalities and differences of Data / Metadata Models from a statistical point of view, irrespective of the terminology and goals of the specific models’. It is based on four views.

1. Conceptual Category View (Conceptual model)
2. Statistical View (Role of the category within the statistical ontology)
3. Data Management View (Access and Manipulation of Category Instance Data)
4. Administration View (Management and bookkeeping of the structures)

This is an important initiative in terms of interoperability and enabling systems to interact by enabling models to be mapped on this framework. It also presents a structure in which the authors of current models that lack a conceptual, or other, framework, can begin to understand their own work in a wider context. In other words it enables them to ‘move up levels’ of understanding and, for example, place the variable label (a variable level might be a particular mode of transport) in the wider knowledge domain. When constructing questionnaires and data browsers there are often many assumptions about categories and themes (again mode of transport or household assets are examples) in which the underlying taxonomy is not explicit, but could benefit from being so. In the public sectors, more sites are using the Government Category List (GCL), part of the interoperability framework. Thisis a starting point and brings more consistency to sites and, potentially, makes it much easier to find relevant information across sites.

DDI model in perspective

The DDI is a structural model for the transfer and storage of surveys and, now, aggregate, data. It takes a document-oriented approach, creating a static codebook which has the aim of describing everything about a set of data that a secondary analysis is expected to need to know. Sections include the study information, a sort of 'Dublin Core plus' approach to describing everything about the background, ownership and methodology of the survey. Then there are sections describing the file and data structure, more analogous to an expanded Triple-S format. Given that it was designed with the conceptual view of data as a static and stand-alone entity, it was not designed with any concept of how that data fits into a wider model or framework. This can be added in terms of a specific domain via the use of keywords, but this depends on external models to make use of them and place the dataset in some wider context. Grossman (2003) has sought to place this metadata container in a wider UMAS model. By placing the DDI structure within a UMAS model structure, it is to be hoped that users would be able to move up the implicit model to higher levels of abstraction. For example, in Grossman analysis, a category text in the DDI could be mapped to a concept dimension in UMAS or ISO 11179 for that matter.

The DDI does not include semantics not least in the fact that there is no link to classification data. Classifications (or value sets) are expected to be repeated in every DDI codebook, even if they are repeated across similar codebooks (just as is the case with SPSS although not SAS). Of course it does not stop the data being modelled properly in a system and the DDI used just as the interchange format, but even then the DDI file would have to include some reference to where those classifications could be found and it is not designed to do that. Studies (or datasets) are the highest level of abstraction – variables live in the context of a study and cannot be abstracted out of that context

Interestingly the guides on data documentation created at the UK and US social science data archives (UK Data Archive (2002) and ICPSR(2002)) demonstrate how much attention has been paid to proper and complete documentation. This is geared towards populating the DDI record. What is now needed is to view that documentation in a wider perspective. As it is created in XML, the DDI does lend itself to being used in web based systems that exploit the XML syntax. As such it is en extremely useful structure and has been utilised by the Nesstar software and others. Given the success of the DDI it is to be hoped that future iterations of the standard will utilise the wider perspective of issues to do with data production, semantics and classification systems.

6. Presentation Issues

Metadata driven

Returning to the end user perspective of our metadata, it is important that the metadata system contains enough information to drive a web browsing, or presentation system. For example a dataset should contain standard geographical identifiers that allows the system to know which map to open in order to create a thematic map interface. If there is a time dimension, that should be identified so that special operations on time can be opened up. Similarly there may be issues to do with disclosure control that require the interface to be altered.

Another good example is the creation of composite tables being driven from multiple data sources. Instead of using a message type system (see GESMES a Eurostat and European Central Bank statistical data exchange message system to exchange statistical data and metadata) to collect disparate

data sources, adequate metadata should allow tables to be built on the fly from distributed data servers. Then, assuming adequate metadata exists, the user is able to drill back from the specific table row or column to find out more about a particular data point.

Persistent urls

Another important presentation issue is persistence. This is the requirement to identify a specific table into the future. There is a significant development around the creation and management of persistent urls (see www.purl.org) and these should be used to create unique identifiers for data and specific manifestations, just as the ISBN number does for books. Other initiatives in the registration of DOIs (Digital Object Identifiers) suggest development sin this area and it is important that the needs of the data analysts, which are specific, are include din these deliberations. For example a pertinent issue is to define what is to be identified, is it a web site with tables, a report with table within it, a table itself, a particular subset of the table or a manifestation of the table (e.g a graphical ot tabular display). My aim is to point out that these issues, and the particular needs of statistical data and metadata are not overlooked.

User centred design

So for the data to flourish it has to be embedded in well-designed metadata driven systems. Firstly we have to ask whether a particular system is useful – in other words has it met the basic user requirements. There is no point creating a highly complex metadata environment when the end user will not be able to understand the information presented. Again, it ought to be possible to use good metadata structures to deliver just the right amount of information depending on various user categories.

It is also necessary to ensure that a system is developed according to good usability principles, in particular by involving the end users at all stages of development. It is notable that many statistical systems are complex and hard to use. With a wider user community the challenge now is to combine advanced usability techniques with systems that provide the level of information consistent with the understanding of the end user. So part of the usability challenge in this context is to ensure that the system designers do more than lock out potentially misleading data or just leave it as unusable, but rather open up channels that the more naïve can follow, in order to understand the content. For example it is often a concern of statisticians that data can be mis-interpreted. Whilst this is self-evidently true, it is not a reason to close access. It is better to focus the contextual metadata on explanation that illuminates the complexity and so allows the user to realise their own limitations. For this to happen, the metadata model must accommodate links to pedagogical material.

7. Examples

As an example of the type of interface driven by the metadata see the screen shot below. In this example, from a system being developed for the World Bank by Nesstar Limited, the table is built up row by column from the categorical variables. The cell content is driven by the measure variable, which is one of the continuous variables, such as population, sample, income etc. In other words the classification of the type of variable (nominal, ordinal or scale) will drive how it is used in the display.

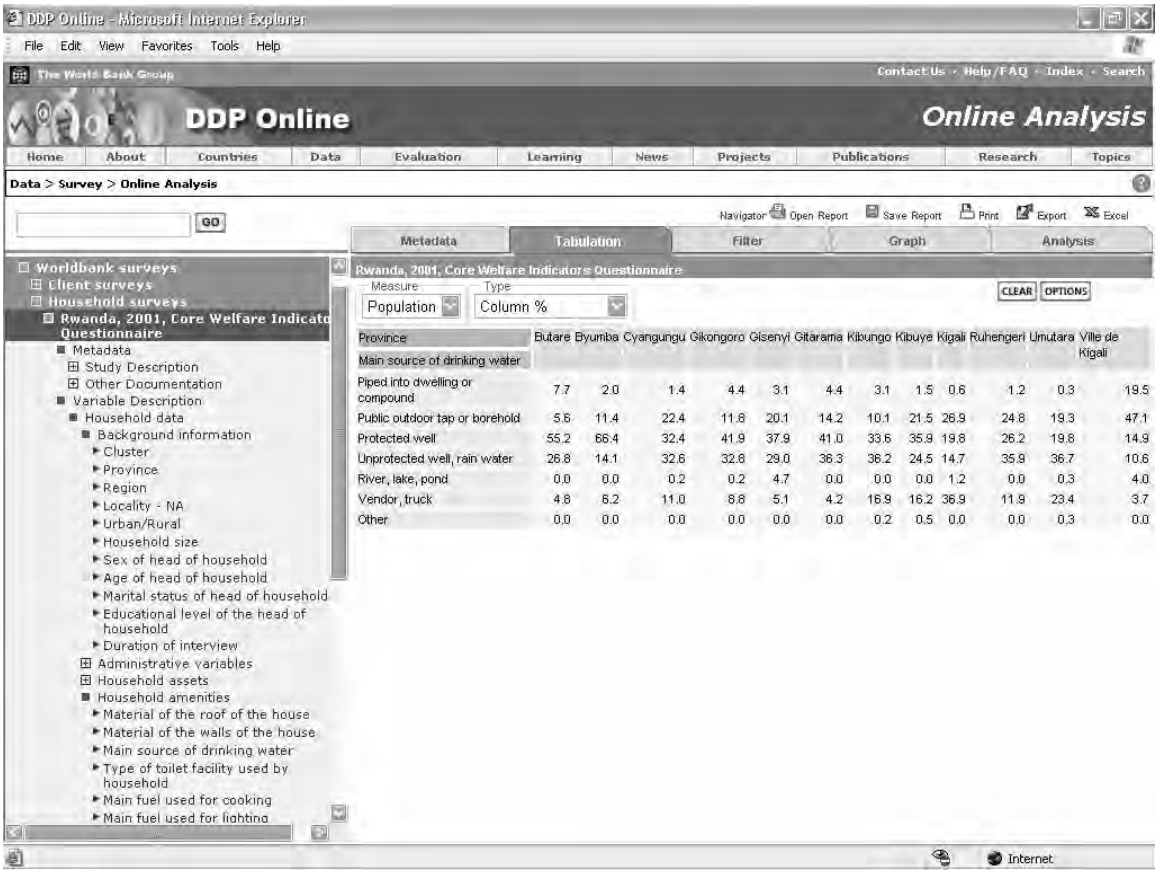


Figure 2: The data analyst's system

An example of a system that incorporates a range of e-GIF compliant metadata is the Black Country observatory below. The browselist on the left shows how the GCL (Government Classification List) is used to group different resources, in this case tables. The frame in the bottom right shows a display of e-GMS metadata, the high level metadata about the table. In the middle right frame is the table itself. This has metadata that cannot be incorporated within e-GMS. Instead the metadata concentrates on structure, dimensions and feasible operations. For example this table has a geographical reference, that defines which map to use, and a time variable, which opens up the time series display.

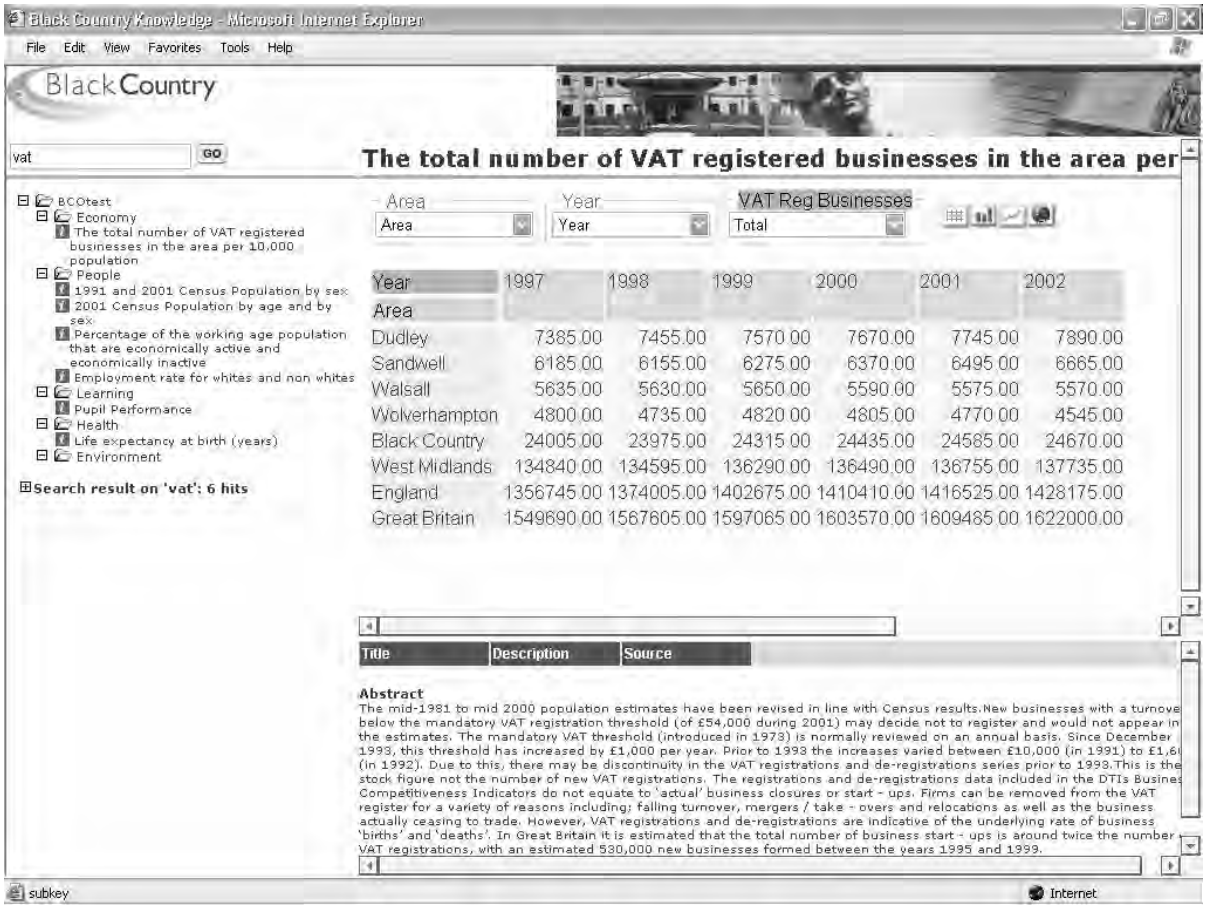


Figure 3: The data observatory

Both of these systems are metadata driven. The first is complex and aimed at the professional user. In contrast the final screen shot shows a more simple interface taken from a data observatory in which key performance tables can be displayed for the more naïve user, but also serve as links back to underlying tables or surveys for the more advanced user.

8. Conclusions

Metadata that is adequately modelled and comprehensive inb scope has the potential to allow users, human and computer, to know where data have come from and where they have the potential to go. As a result, in today's semantic web environment, data can be freed from static manifestations or exchange formats to become increasingly flexible and powerful by integrating with a wider range of complementary resources in the knowledge economy. Interestingly the increased potential usage of statistics places a growing emphasis on the pedagogical nature of the metadata and information structures generally.

References

Assini P (2002) NESSTAR: A Semantic Web Application for Statistical Data and Metadata, in Real World Semantic Web Applications, Vipul Kashyap and Leon Shklar eds. IOS Press, Amsterdam, 2002, ISBN 1 58603 306 9, pages. 173-183.

Eastern Region Public Health Observatory: Public Health Information Tagging System, version 1.2.1 (2003) www.phits.org.uk

Froeschl K A, Grossmann W and Del Vecchio V (2003) The Concept of Statistical Metadata in MetaNet Work Group 2 Harmonization of Metadata: Structure and Definitions

ICPSR (2002): Guide to Social Science Data Preparation and Archiving at www.icpsr.umich.edu/access/dpm.html

GESMES See www.ecb.int/stats/gesmes/gesmes.htm

Grossmann W (2003) Presentation at Metanet Conference INEAG, Samos, May 2003

Hendler J, Berners-Lee T and Lassila O the Semantic Web in Scientific American, May 2001

Ryssevik J and Musgrave S (2001) The Social Science Dream Machine: Resource Discovery, Analysis, and Delivery on the Web, in *Social Science Computer Review* Volume 19, No. 2, Summer, 2001

Ryssevik J (2000) Bazaar Style Metadata In The Age Of The Web - An 'Open Source' Approach To Metadata Development, in UN/ECE Work Session on Statistical Metadata (Washington D.C., United States, 28-30 November 2000)

Local Futures: see www.localfutures.com/strategy/audits.asp

Triple-S: www.triple-s.org/ssssgc92.htm

UK Data Archive (2002) Good Practice in Data Documentation at www.data-archive.ac.uk/creatingData/goodPractice.pdf

UK Government (2003) e-GIF Version 5 Part 2 http://www.govtalk.gov.uk/

Westlake A (2003) Presentation at Metanet Conference INEAG, Samos, May 2003

About the Author

Simon Musgrave is currently marketing director at Nesstar Limited a company set up to exploit the Nesstar software. Prior to the foundation of Nesstar, Simon was Deputy Director of The Data Archive, at the University of Essex. Simon was leader of the original NESSTAR and FASTER EC-funded projects that developed semantic web based data discovery and analysis software. He is an expert on statistical data dissemination and has been involved in numerous development, implementation and consultancy projects, working, amongst others, with data archives, health departments, regional observatories, international development agencies and market research. He can be contacted at Nesstar Limited, John Tabor House, Wivenhoe Park, Colchester CO4 3SQ; tel. 01206 872321; e-mail simon@essex.ac.uk

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

The OnQ Survey Database System. Architecture and Implementation

James C. Witte & Roy P. Pargas

Abstract

This paper discusses the architecture and implementation of *OnQ*, a web-based survey database system. The *OnQ* system is an outgrowth of a US National Science Foundation funded project, *Survey2001*In this paper we specifically focus on several behind the scenes features of the *OnQ* system including the database architecture, and the presentation manager and the *OnQ* authoring tool.

Keywords

Web surveys, survey methods, survey authoring tool

1. Introduction

As part of their coverage of the millennium, the National Geographic Society collaborated with social science researchers to use a new research tool, a web-based survey, to tackle an age-old question, "How does where you live shape who you are?" Launched in October 1998, *Survey2000* was an ambitious experiment in web survey methodology and technology. The survey was developed on a Unix workstation using CGI scripts. Questions, answers, and the ability to proceed to one question versus another, as determined by a given question's answer, were all hard coded in the CGI scripts (Jaganathan 2002; Pargas et al. 2003b). *Survey2000* asked how often people moved, how strongly they felt about their communities, how extensively they used the Internet, and measured their global and local cultural preferences in terms of food, music, or authors. Magazine and television advertisements were used to reach potential respondents and more than 80,000 respondents in over 175 countries participated in the survey over a ten-week period (Witte et al. 2000).

Based on the success of *Survey2000*, a follow-up project was funded by the U.S. National Science Foundation (NSF). Once again with the cooperation of the National Geographic Society, *Survey2001* studied the impact information technology, particularly the Internet, has had on contemporary society. The survey focussed on impacts in three areas of society: community, culture and conservation. The overarching substantive goal of the survey was to use these three areas to describe how the Internet redefines our sense of geography, particularly the distinction between the global and the local. Launched in October 2001, the survey was available in four languages, English, Spanish, German, and Italian (Pargas et al. 2003b; Witte 2003) and collected responses through September 2002.

One key feature of the project was the *Survey2001* database design, which allowed for a dynamic online questionnaire in an open Internet environment. The dynamic character of *Survey2001* was introduced in the opening screen, where respondents select to take the survey in one of the four

languages. Similarly, on the following screen respondents were steered to an adult or a youth version of the survey depending on their age. Next, respondents supplied additional demographic information, including current primary residence, marital status and household composition. Subsequent questions asked about race and ethnicity, educational enrolment and attainment and current employment status. This initial section was very important because many subsequent questions were filtered or tailored based on the respondents' demographic data. These items came from a variety of topical modules including Internet access and use, environmental issues (both global and local to where the respondent lived), community, reading behaviour, politics, science knowledge, and cultural tastes and preferences.

Even with its database foundation, implementing *Survey2001* was quite labour intensive. The survey questions and possible answers were entered into the database, to a large extent, by hand, and the effort required to accurately insert over 2,000 questions and 1,500 answer options (for each language), was high. Some errors took hours to locate and every substantive or editorial change required a database update. The reliance on the data entry person, and the need for meticulous precision, was extremely high. Furthermore, different sections of the questions of the survey were written by different sociologists (in part because the survey was to be duplicated in several different languages), but the questions and answer options could ultimately only be entered by computer scientists familiar with database SQL statements. During the development of the German version of the survey, for example, German translators received copies of English questions, translated them into German, and returned the translated questions in an Excel spreadsheet. Computer scientists then wrote programs that read from the spreadsheets and inserted the questions into the database. The German translators then checked the accuracy of the survey content (on the Web), and recommended changes and corrections, which had to be made manually (again by a computer scientist). Entering, checking, and testing of the entire survey took hundreds of hours. In total, it took eight sociologists, three computer scientists, and six University and National Geographic personnel over six months to enter and edit the data and develop the software to create *Survey2001*.

So, as an extension of the *Survey2001* project the *OnQ* authoring tool, was developed. The *OnQ* authoring tool is a Java-based application with a graphic user interface (GUI) that provides the means for a survey author to insert questions, answer options, skip patterns and formatting rules. As a result the survey author can create a well-formatted, dynamic web-based survey comparable to *Survey2001* without needing to know Java, MySQL or even HTML (Pargas et al. 2003a).

For a variety of reasons, *Survey2001* provided an ideal opportunity for developing web survey tools. Specifically, *Survey2001* (a) was multi-lingual, (b) had one version for minors (under eighteen) and another for adults, (c) required moderately complex skip patterns (transitions) from one block of questions to others based on responses to multiple questions, (d) required images to accompany some of the questions for half of the respondents and no images for the other half, (e) had four mandatory and four optional categories of questions, (f) presented optional categories in random order, (g) allowed the respondent to enter comments at any time, (h) recorded what questions the client was viewing when the comment was made, (i) provided visual queues indicating progress made through the survey, and (j) worked with an external data database (of zip codes and the corresponding cities and states) in developing the text for one of the questions.

In short, if the tools were able to meet the project requirements of *Survey2001*, then they were likely to be more than adequate for most survey situations. The following sections explain how these requirements were addressed by a combination of database design, and presentation and authoring software.

2. Architecture

An overview of software and the database supporting *Survey2001* is shown in Figure 1. The database system used is MySQL (version 3.23.47). The survey is deployed by a collection of servlets, called the *Presentation Manager*, whose HTML output is served to the client by an Apache web server and Tomcat servlet container (version 3.2.3). The database tables are divided into five general groups, supporting (a) questions and answers, (b) question blocks, (c) responses and comments, (d) presentation format and layout, and (e) overall survey information. Details on each are provided below.

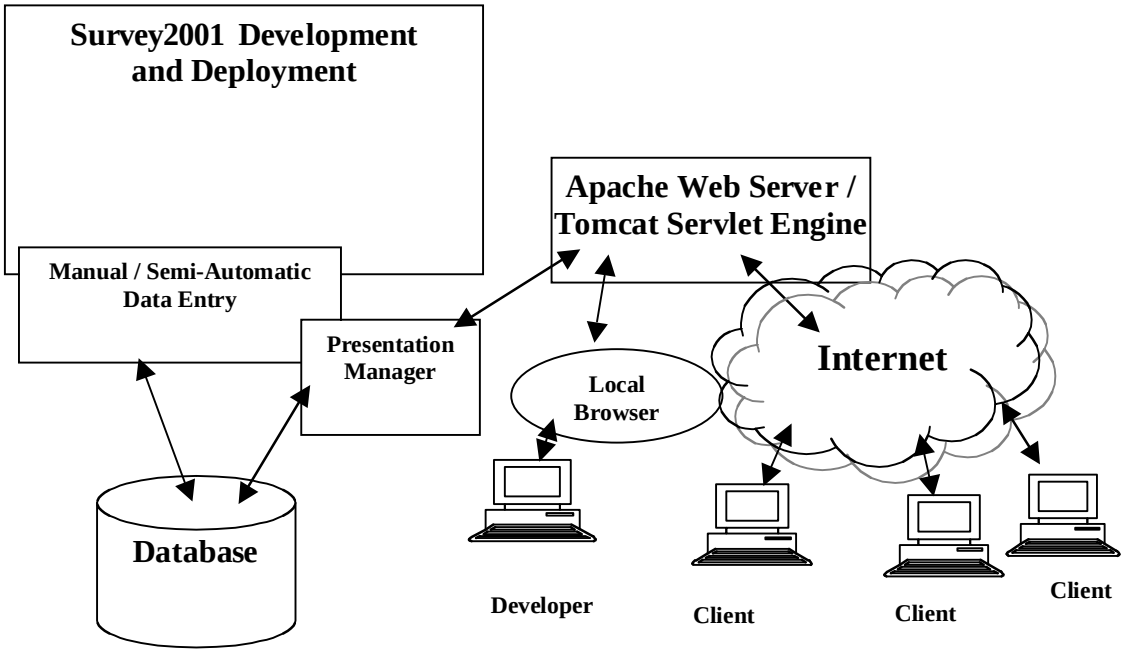


Figure 1. **Overview of *Survey2001*: development and deployment.**

Questions and Answers

Consider a survey that contains the following items.

Rate each of the following foods

	Awful	So-so	Good	Great
Hamburgers	O	O	O	O
Soft tofu	O	O	O	O
Pizza	O	O	O	O
Fried chicken	O	O	O	O
Bean sprouts	O	O	O	O

which have been suggested for the office picnic lunch.

In the database, this example is represented by five questions whose `root` fields contain the strings "Hamburgers", "Soft tofu", "Pizza", "Fried chicken", and "Bean sprouts". The `answer_type` for each question is "radio button" and each question has four `answer_options`: "Awful", "So-so", "Good", and "Great". Each question is also associated with records in the `prefix` and `suffix` tables, which hold the fragments "Rate each of the following foods" and "which have been suggested for the office

picnic lunch," respectively. Finally, the questions are grouped into a single radio button table with `answer_options` aligned in columns.

The same five questions can be presented in different ways based on information also stored in the table. The questions may be presented individually rather than in a table. And if so, the `answer_type` for one or more of the questions may be a "drop-down menu" instead. Each question may have a different prefix and suffix, may have up to three media components (still images, audio, or video), and may be presented in one of several different formats. Or all five questions may be completely reorganized as one multiple-response, check box question in which the client clicks on all acceptable choices ("Hamburgers", etc.).

Question Blocks and Transitions

Questions are grouped into categories (for example, some of the categories in *Survey2001* were *Demography*, *Internet*, *Conservation*, and *Culture*). Within a category, questions are grouped into question blocks. For now we can think of a question block as a set of questions. A more formal definition of the rules defining such sets is found below.

The concept of a question block, and the ability of the survey author to specify that transitions, or transfer of control, conditionally take place from one block to another, provides the dynamism in online surveys. Figure 2 shows an example of four question blocks (QB1, QB4, QB25, and END) and transitions from one to another. In the example, control transfers from QB1 to QB4 if Boolean expression T1 evaluates TRUE. Transitions are arbitrary Boolean expressions whose variables are questions that the client has already seen and answered. For example, transition T1 from QB1 may read:

> *Q5=A10 and Q7=A9*

which, in turn, is interpreted by the *Presentation Manager* as:

> *From block QB1, move to block QB4 if the client answered* 'Yes' *(i.e., answer A10) to question Q5 and* 'No' *(i.e., answer A9) to question Q7.*

If the expression is FALSE, then control will transfer from QB1 to another question block (not shown).

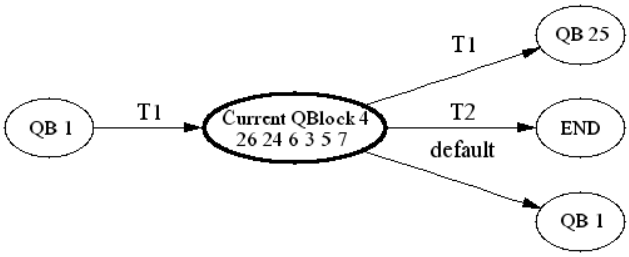


Figure 2. ***OnQ* graphic view of question blocks and transitions**

From QB4, control transfers to QB25 if T1 evaluates TRUE, else the survey ends if T2 evaluates TRUE. Otherwise control transfers back to QB1. Note that transition T1 from QB1 is *not* the same as transition T1 from QB4 and that the default transition from any node *always* evaluates TRUE. Boolean expressions are evaluated in increasing index order and the default expression is evaluated last.

This simple mechanism is a powerful tool. The author may now view the entire survey as a directed graph representing a deterministic finite state automaton where nodes are questions blocks and

transitions are controlled by Boolean expressions. An example is shown in Figure 3. Two distinguished blocks, void of questions and called START and END, mark the start and end of the survey. Initially, START is designated as current; at any other time, the block whose questions the client is answering at the present time is considered *current*. The client's answers determine which Boolean expressions will evaluate TRUE. These, in turn, determine which block is selected next.

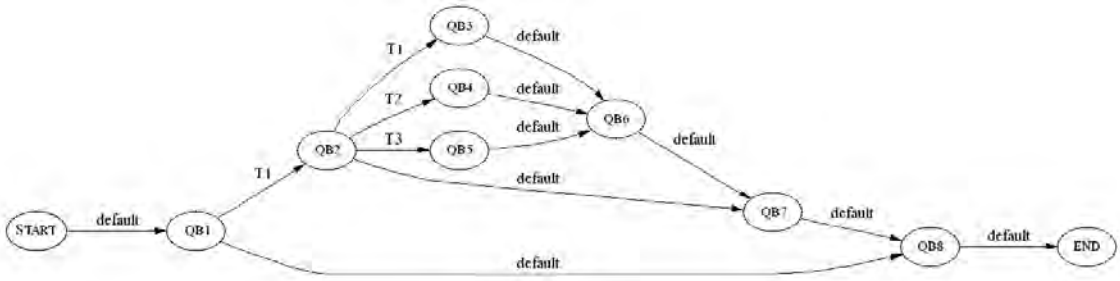


Figure 3. ***OnQ* graphic view of an entire survey.**

The two tables used to support transitioning are shown in Figure 4. Table `question_block` assigns an ID number to each block. Questions contained in the block are listed in field `block_seq`. Fields `rand_paramter` and `rand_type` deal with whether the questions in the block are to be randomised before being presented to the client. Table `block_sequence` provides 5-tuple records of the form:

```
(survey_type,current_block,expression,next_block,seqnum)
```

Table question_block;

Field	Type	Null	Key
q_block_id	bigint(20)		PRI
q_block_seq	text	YES	
rand_type	bigint(20)	YES	
rand_parameter	bigint(20)	YES	

Table block_sequence;

Field	Type	Null	Key
survey_type	bigint(20)		
current_block	bigint(20)		PRI
expression	text	YES	
next_block	bigint(20)		
seq_num	bigint(20)		PRI

Figure 4. Tables supporting question blocking and transitioning.

enabling the *Presentation Manager* to decide which block of questions to present next. Expressions are evaluated in `seqnum` order and the `next_block` whose `expression` evaluates TRUE first is selected. The field `survey_type_id` allows the survey author to develop different graphs for different survey types. For example, a similar set of questions, possibly blocked and sequenced differently, may be used in a telephone survey. A telephone operator may click through the survey as he or she asks the respondent the questions. Or a survey taker in a public area, such as a shopping mall, may use a handheld digital device with a wireless connection to the Internet may use an abbreviated version of a survey to assess public opinion.

Responses and Comments

Client responses are kept in two tables, one for fixed answer choices (Table: `user_response`) and another for open-ended answers (Table: `user_response_text`). Both record the client identification (`user_id`) and the question number (`q_id`). In both, answers are date- and time-stamped (`r_datetime`), providing the survey developer with an answer trail for clients who click the back button on their browser and answer questions multiple times. They differ only in the answer field. Table `user_response` stores an index identifying the selected answer option (`a_id`) into the table of answer choices. Table `user_response_text` records the client's text answer. In the event

that the user has used the back button and has responded multiple times the respondent's final answer based on the date and time stamp is used to govern future sequencing between question blocks.

At any time during the survey, the client may submit a comment by clicking on a "Submit a Comment" button on each page. Comments are stored in their own table (Table: `comments`) and are date- and time-stamped. The current question block number is stored in a field `q_block_id` in order to provide the survey developer with the context within which the client made the comment. However, the number of questions within the question block may be large. To narrow the context down, field `questions` lists precisely those questions that the client has on his or her screen when the comment is submitted.

Presentation Format and Layout

The survey developer can provide the *Presentation Manager* with formatting requests for individual questions and for blocks of questions. For example, say that question block 1 of survey type 1 contains questions 1 through 10, but the developer wants to present to the client first three, then five, and finally two questions on three separate web pages. Entries of (3,1,1,"EP"), (8,1,1,"EP") in Table `presentation_format` (Figure 5) tells the *Presentation Manager* to end a web page ("EP" stands for "end page") with questions 3 and 8. The final two questions of the block, i.e., questions 9 and 10, will be presented on a separate page because the *Presentation Manager* will end a web page automatically when a question block is exhausted.

```
Table presentation_format;
+---------------------+------------+------+----+
|Field                | Type       | Null | Key|
+---------------------+------------+------+----+
|q_id                 | bigint(20) |      | PRI|
|q_block_id           | bigint(20) |      | PRI|
|survey_type_id       | bigint(20) |      | PRI|
|q_presentation_format| text       | YES  |    |
+---------------------+------------+------+----+
```

Figure 5. Question and block formatting information

If, further, the survey developer wants the five questions on the second page to form a *radio button table*, `presentation_format` entries of (4,1,1,"RBT1"), (5,1,1,"RBT1"), …, (8,1,1,"RBT1 EP") inform the *Presentation Manager* that questions 4 through 8 are to be presented collectively in a *radio button table*. The field `q_presentation_format` also enables the author to select from several question layouts, which determine the relative positions of questions and answers on a page.

Overall Survey Information

The *OnQ* database table, called `survey_instance`, contains general information about the client. This information includes the language in which the client chose to take the survey, the time the client started, information about the client's computing environment (for example, browser name and version), the URL with which the client accessed the survey, whether the client is a referral by an earlier respondent, and if so, the user id of the referring client.

The URL from which the client linked to the survey is of particular interest to the survey analyst because this may provide insight into potential biases common among clients who frequent that URL. For example, it may be informative to group responses from clients who linked to a survey from a politically moderate website and contrast them to responses from clients who connected from liberal and conservative websites. This may assist the analyst in more accurately assessing sources of sample bias.

3. The *OnQ* Authoring Tool

After developing *Survey2001*, it was clear that the need for a survey-authoring tool was great. The problem at the time was that none of the authoring software available when *Survey2001* was being developed satisfied all of the requirements of *Survey2001*. Some of the authoring tools considered are described below and a table comparing them to *OnQ* is displayed in Figure 6.

Web Surveyor (http://www.websurveyor.com/home.asp) is available for PCs running Win95 and later operating systems, or it can be used from the WebSurveyor server. The system can handle surveys, versatile branching, different background colours, and different text options.

WWW Survey Assistant (http://or.psychology.dal.ca/~wcs/hidden/home.html) is Web-accessible, and works with Win95 or later. Because it is Java-based, it can be run on any machine, including Solaris, OS/2, and Macs. All project files are saved on their server and provides multiple text fonts. The software claims to randomise questions and to handle complex skip patterns but details are sketchy.

Hosted Survey (http://www.hostedsurvey.com/Spotlight.htm) is entirely web-based. It handles conditional branching where the answer to a single question can branch to a series of questions, and it can randomise questions. The tool allows questions to be grouped into sections.

Survey Said for the Web (http://www.surveysaidfortheweb.com) is Web-accessible and works with Win95 or later. This tool allows for single, linear skips only. One feature of this authoring tool is the use of keyword association for easily identifying questions, and also a search feature for the questions where it can search by word or phrase.

Finally, *Sawtooth Sensus Web* (http://www.sawtooth.com) is Web-accessible and works with Win95 or later. This tool handles branches of any complexity, and can randomise the answers, questions, or blocks of questions.

	Web Survey	**Survey Assistant**	**Hosted Survey**	**Survey Said on the Web**	**Sawtooth Sensus Web**	***OnQ***
Win95+	*	*		*	*	*
Other machines		*				*
Images & audio	*	*		*	*	*
Question categories			*			*
Branching	*	*	*	*	*	*
Randomisation		*			*	*
Multi-language	*					*
Keyword association				*		*
Different backgrounds	*					*
Different text type	*	*				*

Figure 6. A summary of web survey authoring tools

Comparing the existing tools and the needs for *Survey2001* set the design requirements for *OnQ*. It was decided that the ideal tool would have a graphical user interface allowing a developer to enter questions and answer options. The program would then write these into the *OnQ* MySQL database, which would then act as the questionnaire backend. This tool would also provide a way for the user to create blocks or groups of questions and add transitions among these blocks (transitions instruct the

survey presentation tool as to which group of questions to present *next* after the current group is answered). Question blocks and transitions would be stored in the database. The developer would also be able to use the tool to define question layout and presentation including grouping questions with identical radio button answers options a compact radio button table. The tool would also allow questions to be randomly presented with a choice of several randomisation functions, including constructed lists. The program would allow access to external databases (such as a zip code table) during the execution of a survey. Finally, such a tool could be run standalone as well as on the web where it could be accessed by clients to develop the survey database.

OnQ Authoring Tool Design Features

The *OnQ* authoring tool is written in Java (version 1.4) because of Java's cross-platform portability as well as Java's suitability to web applications. *OnQ* accesses a MySQL database (ver. 3.23.47) using a MySQL Connector/J Driver. Pages are served using Apache as the primary web server and the Tomcat (version 3.2.3) from the Apache Group as the servlet engine.

Architecture

The design of *OnQ* is purposely organized around the steps we anticipate a typical user will take to develop a survey. When *OnQ* starts, the survey author enters a user name and password, and chooses which module and which database to use. If desired, a new database may be created by entering a name for the new database.

With a new database, the survey author will first want to work with the *Question Editor*, which enables the author to enter and modify questions and answers. The author selects answer option formats (radio buttons, drop-down menus, checkboxes, textboxes, or text areas) and media components (image, audio, video, or animation). The *Question Editor* allows the association of a key with each question to facilitate searching for a question at a later time and provides a mechanism for the author to group questions into categories, enabling him or her to work on one related set of questions at a time (Hochrine 2002).

The next step is to form logical subgroups called *question blocks*. This function is provided by the *Sequence Developer* that enables the author to specify the presentation order for individual questions as well as question blocks (i.e., by creating Boolean expression-controlled transitions between pairs of blocks). This is explained in greater detail below. The author may also specify that questions within a question block are to be presented in some random fashion. Error checks within this module eliminates the possibility that the *Presentation Manager* will have to deal with errors such as incorrect transitions or question blocks, or questions not included in the survey.

After the creation of question blocks and transition sequencing rules, the author can use the *Question and Block Formatter* module to specify the layout of a question or of a question block. By default all questions within a question block are presented on a single web page when the survey is deployed. The *Question Formatter* provides the developer with a number of question layout options. For each question, media components may be placed above, beside, or below the question text, question prefixes and suffixes may be joined to or separated from the question root. Using the *Block Formatter*, the developer may choose to present a large question block to the client in a series of pages with a small number of questions on each page. This module also provides an option to display a sequence of radio button questions as a single radio button table.

The interaction among the modules (*Question Editor*, *Sequence Developer*, and *Question and Block Formatters*) is purposely kept simple and limited. Each module writes data directly to the database as changes are made and other modules update their data from the database each time they are accessed. A user can access any of the modules in any order through tabs at the top of the program frame. However, the *Question Editor* must be used to enter at least one question and one set of answers before the other modules can have some data with which to operate.

After questions and answers are entered, images are assigned, question blocks and block sequencing rules are defined (transitions), and the questions are formatted, the developer can use the fourth component of *OnQ*, the *Presentation Manager*, to view the completed survey. The *Presentation Manager* is a Java servlet that extracts the survey information directly from the database and uses this information to serve survey web pages. All of the information it needs is stored in the database. The *Presentation Manager* presents questions, answers, and accompanying media components entered with the *Question Editor*, in the order specified by the author in the *Sequence Developer*, and according to the question and block formats specified in the *Formatter* module. During development of the survey, the author's *local* browser can present the pages exactly as they would be presented on the Web. Additions and corrections can be made directly to the database using *OnQ* and the author can view the changes immediately. The author then deploys the survey on the Web when he or she is satisfied that the survey is complete and correct.

Question Blocks

The concept of question blocks is central to the correct presentation of a survey and requires more explanation. Certain policies were set for questions within a block. First, it was decided that (a) if a question block is not randomised, then every question within a block is to be presented in the author-specified order, before the survey moves to another block; and (b) if the questions within a block *are* randomised in some way, then the questions selected by the randomisation function are presented in the order specified by the function, before the survey moves to another block. Second, adding a question to a question block involves placing a question in sequence within that question block. And third, question block sequencing (as defined by question block transitions) determines the order in which question blocks are presented in the overall survey.

To implement this, the *Sequence Developer* initially creates two blocks called START and END. Both blocks are empty (i.e., contain no questions) and START, by default, points to END. The author must change the default transition from START to a non-empty question block (otherwise the survey will begin and end without asking any questions). This non-empty block contains the first questions to be presented in the survey.

From the first non-empty question block, the author can specify the *conditions* under which control passes to other question blocks. These conditions are defined by an author defined Boolean expression and may be as complex as the author chooses. It is assumed that all question variables used in the expression represent questions already presented to the client. Otherwise, the expression evaluator of the *Presentation Manager* will report that the expression cannot be evaluated and the survey cannot proceed.

To model transitioning among blocks, we use a deterministic finite state automaton or accepter (DFSA) described as follows:

$$M = (Q, \Sigma, \delta, q_0, F)$$

where

M = the automaton and
Q = set of internal states,
Σ = set of symbols (input alphabet),
q_0 = initial state,
F = set of final states, and
$\delta = Q \times \Sigma \rightarrow Q$

(δ is a total function called the transition function which gives the next state in terms of current state or configuration and input symbol; a move is a transition by the automaton from one configuration to another).

The *Presentation Manager* interprets the information recorded in the database in such a way that it essentially moves though the automaton by evaluating Boolean expressions and transitioning to target question blocks accordingly.

One difference, though, is that in a traditional DFSA only one transition (one edge) can be taken based on a given input. In the *OnQ* model there may be multiple transitions that evaluate to TRUE when the survey is taken. Each transition from a question block is numbered from 1 upward. Transitions from one question block are evaluated in order until the first expression evaluating to TRUE is found, and the corresponding transition to a next block is taken. Also, the last transition must be a default transition that *always* evaluates TRUE, guaranteeing that a transition will be taken.

Graphical Representation of Question Blocks

Transition graphs are traditionally used to visualise and represent finite state automata. In a transition graph, vertices represent states, edges represent transitions, labels on vertices represent names of states, labels on edges represent current values of input symbol. For example an edge labelled with (q_0, q_1) represents the transition function $\delta(q_0,a)=q_1$ (usually read as, "from state 0 with the input of an 'a,' move to state 1") (Linz 1996).

A useful feature of *OnQ* is that it can present to the author a graphical view either of a single question block (Figure 2) with all transitions to and from the block, or of the entire survey (Figure 3). To implement this, a software tool called *dot* contained in a package of graph drawing tools called *Graphviz* from AT&T Research is used. *Dot* is a pre-processor which reads a script file description of a graph and draws the graph in one of several formats, including GIF, JPG, and PS. The script and graph are developed by *OnQ* and by *dot*, respectively, in real-time, reflecting all changes made by the author to question blocks and transitions up to that time. To our knowledge, no other online survey authoring software offers anything analogous to this.

Error Checking

Using the question block model of function also allowed us to develop some algorithms to check user input for possible errors. One function the *Sequence Developer* provides is alerting the user if a question is not a member of any question block or a question block is not part of the survey's network of blocks and transitions.

OnQ in Action

Our initial goals were two-fold: (a) to develop a database design for dynamic online surveys, and (b) to develop an authoring tool powerful enough for an inexpert computer user to develop and deploy dynamic and complex surveys. Through the use of the OnQ authoring tool and database, a straightforward process of web survey development and deployment has emerged. Developing a new

survey requires duplicating the generic database structure and using *OnQ* to enter new questions, answers and transition rules into the appropriate tables. Launching a new survey is equally straightforward; new graphic images are created giving each survey a distinctive look and feel and masking the fact that the essential operation of the presentation manager is unchanged.

Overall, *OnQ* fulfils nearly all of our original goals for a survey-authoring tool. First, *OnQ* creates and populates all the necessary database tables with enough data to present a large and dynamic survey. Second, we feel that *OnQ* is easy to use. *OnQ* offers non-technical users the simplicity of a point-and-click interface. Third, the *OnQ* authoring tool has already been used to develop several online surveys.

- Diversity2002 is the second in a series of four annual surveys of the Clemson University Class of 2005. This four-year study hopes to come up with recommendations on how to improve campus life by reducing or eliminating situations that may create racial tensions.
- UnitedWay2002 collected opinions and ideas from members of the Greenville, SC community regarding local community strengths and issues. This effort is led by the United Way of Greenville County, SC with partners representing various sectors of the community.
- IE2002 was conducted by the Clemson University Industrial Engineering Department to survey industrial engineers around the country and compare Industrial Engineering curricula. Part of a National Science Foundation funded initiative the primary objective is "the development of a new scaleable and deployable industrial engineering baccalaureate-degree renewal model."
- CCI2003 is a survey of guests and clients who use the Conference Center and Inn at Clemson University. The aim of this survey is to identify strengths and weakness of the guest facilities.
- CUService is a survey initiated by the Office of the President at Clemson University. Based on the premise that universities deliver a wide range of services to different constituencies (alumni, students, faculty and staff) the survey queries respondents regarding their service experiences as delivered by over forty different on campus units.

4. Summary and Future Work

The database described in this paper has served us well over the past one-and-a-half years, and continues to do so. New surveys with new content and completely different skip patterns are now developed quite easily. Each survey, however, suggests new ideas on how to make *OnQ* more powerful and our goal is to continue improving the database, modifying tables as new needs arise. For the immediate future, changes to the database will: (a) allow the author to specify type fonts and sizes of prefixes, roots, and suffixes, (b) enable the author to specify different backgrounds for different categories, (c) add a new question type, the email invitation question, that causes the *Presentation Manager* to send email invitations (to take the survey) to addresses specified by the client, and (d) print paper equivalents (with appropriate skip instructions) of an online survey.

Second, a fully functional, web-version of the authoring tool is planned. The developer will be able to click on a URL and access the authoring tool remotely. The database will be stored, by default, on the server on which web-*OnQ* resides. The developer decides when to release the survey for public use. Until that time, only the author can modify and view the survey being developed.

Third, a DVD training program is planned for development. The DVD will accompany the software and will help a novice user work through the survey authoring process including instructions in the proper construction of survey questions, pros and cons of the different answer options, and providing a large number of examples on how to use the various features of *OnQ*.

Finally, there are plans to develop *OnX*, short for *Online eXercises/eXamination*. *OnX* is the next step beyond *OnQ*, and uses *OnQ*'s architecture and database design. *OnX* will be an authoring tool that allows a teacher to develop web-based learning and assessment instruments. The overall goal of *OnX* is to help pre-K through 12th grade teachers develop software tools and a strategy for coordinating and delivering standards-based curricula.

References

Hochrine C (2002) *The Development of the Question Editor for OnQ, a Survey Authoring Tool*. M.S. research paper. Department of Computer Science, Clemson University, Clemson, SC.

Jaganathan K (2002) *Dynamic Web based Survey: Architecture and Implementation*. M.S. research paper. Department. of Computer Science, Clemson University, Clemson, SC.

Linz P (1996) *An Introduction to Formal Languages and Automata*, 2nd edition. D.C. Heath and Company, Lexington, Massachusetts.

Pargas RP, Witte J C, Brand L, Hochrine C and Staton M. (2003a). *OnQ: An Authoring Tool for Dynamic Online Surveys*, ITCC 2003, Las Vegas, NV, April 28-30, 2003, pp. 717-723.

Pargas R P, Witte JC, Jaganathan K, and Davis J S (2003b) *Database Design for Online Dynamic Surveys: A Case Study*. ITCC 2003, Las Vegas, NV, April 28-30, 2003, pp. 665-671.

Witte J C (2003) "Survey Methods - Large Sample Design & Studying Life Online", *The Internet & American Life*, Philip E.N. Howard and Steven Jones (eds.), Thousand Oaks, CA: Sage Publications.

Witte J C and Howard P E N (2002) Technological & Methodological Innovation in Survey Instruments: The Future of Polling." 2002. (James Witte and Phillip Howard). In: *Navigating Public Opinion*. Oxford University Press. (ed. Fay Cook and Jeff Manza)

Witte JC , Amoroso L and Howard P E N (2000) "Method and Representation in Internet-Based Survey Tools: Mobility, Community, and Cultural Identity in Survey2000," *Social Science Computer Review*, Summer 2000 Special Issue, 18(2).

About the Authors

James Witte is an Associate Professor of Sociology and Director of Survey Methods and Operations for the Center for Advancement of Marketing and Social Science (CAMSS) at Clemson University, Clemson, SC, USA. He was the principal investigator for *Survey2000* and *Survey2001*. He regularly teaches courses on survey design, research methods and the sociology of work. He can be contacted at Department of Sociology, Clemson University, Clemson, SC 29634-1354; tel. 864.656.3816; e-mail jwitte@clemson.edu.

Roy Pargas is an Associate Professor of Computer Science at Clemson University, Clemson, SC, USA. He is primarily responsible for leading the effort in developing the software package *OnQ* which was first used to develop and deploy *Survey2001*. He regularly teaches courses on design and development of network applications. Topics in these courses include Java servlets, Java Server Pages, Java Database Connectivity, XML and XSL. He is a member of ACM, IEEE, and SPIE. He can be contacted at Department of Computer Science, Clemson University, Clemson, SC 29634-0974; tel. 864.656.5855; e-mail pargas@cs.clemson.edu.

Support for this work has been provided by the National Science Foundation (NSF-ITR/Soc (Award # 0082750) and Clemson University.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Developing a Software Suite for Intelligent Questionnaires. Interoperability and Standards

Defeng Ma, Joanne Lamb, Adam Taylor

Abstract

The IQML project automates and integrates the data collection process by producing software to support and implement emerging metadata standards using definition tools such as XML. This paper will present the system structure of the IQML system and the metadata models used. It will describe how the interoperability of the different standards has been achieved in the project. Based on the evaluation and feedback of the users in the trials, it will also discuss the experiences and implications of the IQML project on interoperability and standards, and how this can impact on the survey process for the future.

Keywords

Questionnaire, Metadata, XML, Interoperability, Standards

1. Introduction

IQML (A Software Suite and Extended Mark-Up Language for Intelligent Questionnaires) is a project funded by the EU under the Framework 5 IST program for 3-years, which finished in April 2003 (IST-1999-10338). The partners consist of three commercial companies (Dimension EDI; Comfact AB; and DESAN Marktonderzoek), two National Statistics Offices (The Central Statistics Office (CSO) of Ireland; and Statistics Norway) and two universities (the National Technical University of Athens; and the University of Edinburgh, which co-ordinated the project).

The ultimate goal of the project is to improve the accuracy and timeliness of statistical data collection from enterprises and individuals, whilst at the same time reducing the burden of statistical reporting on enterprises. It also aims to research the realities of metadata interchange and object standards in order to facilitate an active contribution to metadata interchange standards by implementing, in software, chosen aspects of the international standards for metadata interchange and by carrying out trials in the area of intelligent questionnaires and interoperability of the standards.

In IQML there are five related software modules (which will be described in more detail later in the paper) for metadata maintenance, questionnaire design, questionnaire presentation, database interrogation and survey administration. Various data models have been developed in the IQML project. A questionnaire and question bank model has been developed for the Questionnaire Design Tool. In an associated project, work is being done on mapping this model to the Data Documentation Initiative (DDI) model to allow metadata exchange between IQML and DDI. The metadata repository

has adopted the ebXML (electronic business eXtensible Mark-up Language) standard. The Questionnaire Presentation Tool XML (eXtensible Mark-up Language) model for on-line questionnaire has been proposed to become a standard (XML4DR). In addition, many existing technologies and standards of object and metadata exchange are used to ensure the interoperability of the model and standards involved in different modules. The five modules share a common data model to enable the interoperability. Certain modules utilise their own model as described before, each of which maps to the common model, but which also contains features that are necessary for their particular functionality.

The project contributes to standards by participating in the development of the Common Warehouse Metadata Metamodel (CWM) of the Object Management Group (OMG). This ensures that emerging models in the domain of object analysis can be used to define the structure, behaviour and visualisation of a statistical questionnaire including its relevant metadata. Furthermore, the resulting Document Type Definition (DTD) and supporting XML software will demonstrate the benefits of both XML and object technology for administrations, enterprises and other organisations in the context of intelligent questionnaires.

The paper will present the common model, the related models, and the mappings between them. It will also describe the activities of the IQML group in disseminating the models to relevant standards groups, in the software development community and in the area of official statistics.

2. What is meant by “Intelligent”?

Intelligence should be used all the way through the questionnaire design process and into the subsequent stages of survey processing. Intelligence, within this process, can take many forms. Within the questionnaire design process there is a great deal of information that is used to realise the goal of producing a questionnaire that is subsequently discarded. This data includes structural, contextual and semantic information as well as validation and navigation rules. However, much of this data can be used as metadata further on in the statistical process for the production of datasets and documentation etc. This is one element of *intelligence* that is inherent in every questionnaire but is very often not used. Another interpretation of intelligence is the tools given to the questionnaire designer, one aspect of which is the re-use of questions stored in question banks. The QDT is a tool for the questionnaire designer with enough intelligence to allow any type of user to design any type of questionnaire at the conceptual level and then realise that same design in many different media[1]. At the same time, it allows the capture of a lot of metadata and the ability to store it so that it can be used, not only by themselves for design of future questionnaires but also in the eventual interpretation of the resulting data. The use of a question bank is important to allow re-use of questions [Brannen, K.(2001)].

The QDT also allows the designer to link the questions to underlying concepts and to the variables representing those concepts. The links can be published into the Metadata Repository as metadata and can then be used for downstreaming processing by other applications.

The innovation in IQML is in the use of a metadata repository, which allows sharing of metadata objects (including question banks) between modules of the IQML system and other software. The access to the Metadata Repository is via a standard web servlet interface, which means that the

[1] The implementation of QDT has been limited to focus on the web/email format due to resource restrictions; however the model has covered the different media.

Metadata Repository can provide access for many users from different locations, to share the same repository [Lamb, J.M.(2001)].

3. IQML modules

The five modules of IQML are follows:

- **QDT** (Questionnaire Design Tool) module enables the user to design and manage questionnaires through the 'question bank' concept, which enables the reuse of questionnaire 'components'.
- **SAT** (Survey Administration Tool) module allows the questionnaires to be integrated with registers and sample frames.
- **Metadata Repository** module supports the storage and retrieval of metadata objects.
- **QPT** (Questionnaire Presentation Tool) module renders the questionnaire for use with web browsers, supports the presentation of questionnaires defined in the QDT and implements validation, navigation and calculation rules.
- **DIT** (Database Interrogation Tool) module supports the extraction of data from popular databases and maps this data to the questionnaire XML.

System architecture

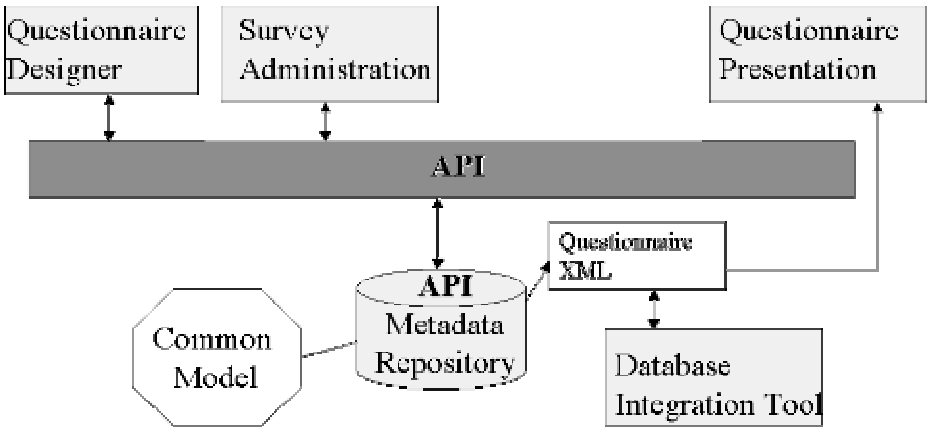


Figure 1. IQML System Architecture

The figure (**Figure 1. IQML System Architecture**) shows the IQML system architecture of the IQML modules. The core module is the Metadata Repository that enables the communication of the other modules.

The QDT and SAT can access the Metadata Repository through the common API (Application Program Interface). Both applications can store metadata to the Repository and can also access the metadata from the Repository.

To communicate with QPT and DIT, a XML file called "Questionnaire XML" can be extracted from the metadata stored in the Metadata repository. Then it will be distributed to the respondents and will be used to collect the response from the respondent through QPT/DIT.

In addition to the five modules described above, the common model and the questionnaire XML are shown in the picture. The common model is used to store the metadata in the Metadata repository (more detailed description later in this paper).

Technical infrastructure and standards

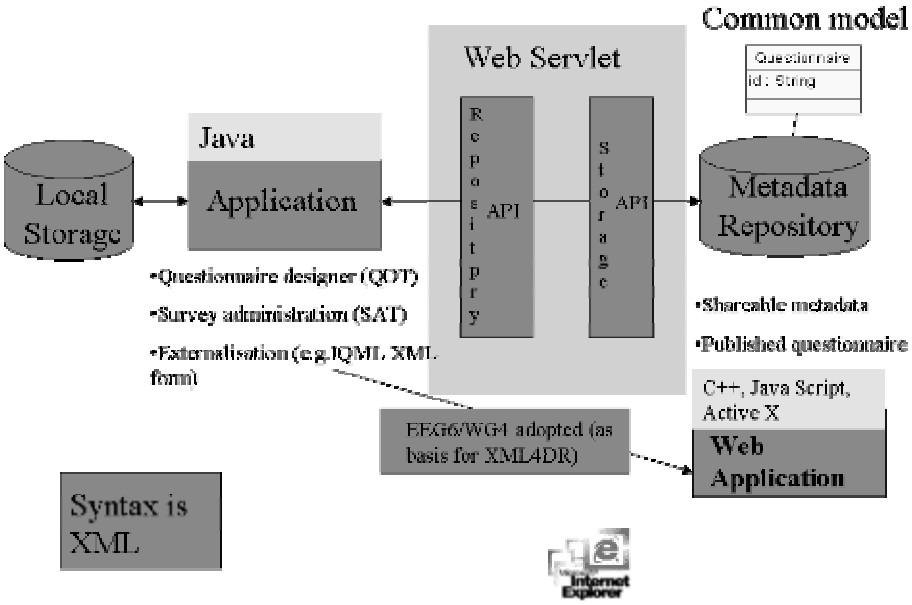


Figure 2. Technical Infrastructure

The figure (**Figure 2. Technical Infrastructure**) shows the technical infrastructure and the standard technology used in the IQML project.

The QDT and SAT are developed with Java technology (J2SDK1.4.0). Both modules can access local storage (a computer hard disk) to store the data. They also access the metadata repository through the HTTP (Hypertext Transfer Protocol) connection to the Metadata Repository's web servlet interface.

The IQML XML file (which contains the questionnaire XML in XML4DR format) can be extracted from the Metadata Repository through the web servlet interface as well.

The QPT is a web application developed using C++, Java Script and Active X. It has used the existing component/functions of the Microsoft Internet Explorer interface. The DIT is developed with C++, Active X and other Microsoft database related technology. The syntax for the exchange of the metadata between the different modules is XML.

QDT

The conceptual basis of the QDT lies in the notion of creating re-usable 'components' as part of the questionnaire design process. In effect they can be thought of as 'building blocks' where elementary components can be combined and recombined to make more complex ones. Having once created them, designers will always have these fundamental components of a questionnaire at their disposal.

Once the questionnaire is ready, it can be published to the metadata repository. All the metadata of the questionnaire, which will be used in the next stage of the survey, will be exported to the Repository by the publishing processing. Other modules of the IQML project or other applications for the downstream processing can use these metadata later.

More detailed description of QDT module can be found in [A. Taylor (2003)].

SAT

The Survey Administration Tool (SAT) is a software module to define, distribute and monitor surveys. It performs sampling, 'prefilling', survey distribution and response monitoring.

The SAT is used to import and display data from a population database; to draw samples - randomly or according to selection criteria - from this population (or sub-samples from a sample) and display these; to create a survey by connecting a questionnaire to one or more samples; to 'prefill' questions in a questionnaire on the basis of information already available in the population database; to add an explanatory message for respondents; to distribute messages and questionnaires by e-mail; to monitor the responses; and to send reminders to all or certain groups of non-respondents.

Based on the monitoring, the SAT can produce the metadata of a survey, such as the response rate.

The Metadata Repository

The Metadata Repository (also called registry/repository) is a shared resource that is used by the other software modules in order to share metadata. Furthermore, it is the intention to store also the metadata that may be needed by downstream processes, such as survey analysis. This shared resource is available to applications using web technology and can therefore be either remote or local to the application. The registry/repository is delivered as two web servlets. One servlet (the "front end") is the application interface and is called by passing an XML file to the URL of the servlet. The other servlet (the "back end") is called by the first servlet and manages the registry/ repository storage.

Therefore, the registry/repository is available to any client application (subject to the usual logging on security) at any location, in a web enabled client/server environment.

QPT

The intended purpose of the Software is to prototype a solution for manual data entry of a Questionnaire. The Questionnaire Presentation Tool is a software product that can interpret an XML document which conforms to the DTD for intelligent questionnaires and to visualise it on a PC. The software allows data entry, performs validations, calculations and navigations as specified in the XML document, and is integrated with popular browser (Internet Explorer[1]) and e-mail systems.

DIT

The intended purpose of the Software is to prototype a solution for automatic data entry of a Questionnaire. The Database Interrogation Tool is a software product that can access an XML document and interface this to databases using standard interfaces in order to automate the link between enterprise data and the questionnaire. It has an interface to popular browsers and email systems.

The DIT supports the semi-automated extraction of data from one or more ODBC (Open Database Compliant) compliant databases. A separate XML file contains the mapping of items in a questionnaire to one or more database. It will save the retrieved data in XML questionnaires. It requires little or no additional software besides a web-browser (IE5+). The DIT is using technologies such as Java Scripts, ODBC, DOM (Domain Object Model), etc. But it requires the knowledge of the database system, location of internal data and the SQL (Structured Query Language).

The data flow in the IQML software suite

Here we describe a typical example of data/process flow in the IQML software suite.

[1] The development of QPT/DIT is based on the IE and other Microsoft technologies. It limits the wide usage of QPT in other Internet browsers.

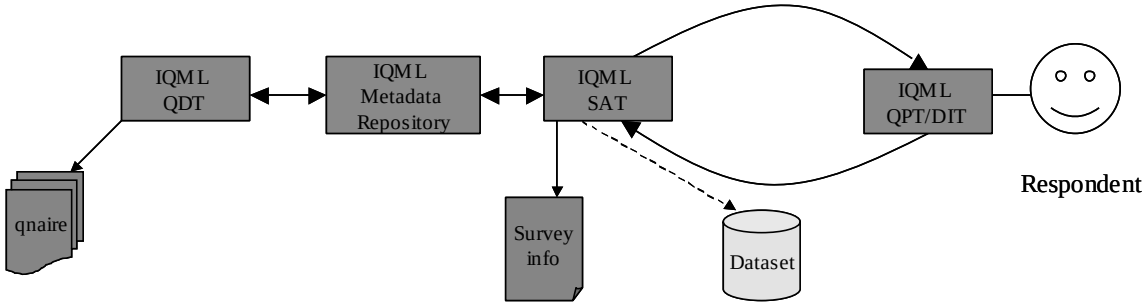


Figure 3. The IQML data flow

The driving process in the IQML project is the survey. The survey manager will draw a sample from the population for the survey using the SAT. Then the SAT can publish the metadata of the population in the Metadata Repository. The metadata of the population is the population information already known by the SAT, like the name, address, contact information etc.

The QDT is used for designing the questionnaire for the survey. Once the questionnaire design is finished, it will be published from QDT to the Metadata repository. During the design stage, a questionnaire designer can view the metadata information of the population from the Metadata repository. To avoid collecting data that already exists (like the name, address), the question can either be omitted from the questionnaire or the question can be defined as a *prefilling* question[1], with the relevant question(s) being prefilled for individual questionnaires with values drawn from the population.

Once the questionnaire for the survey is ready in the Metadata repository, the SAT can view all the questionnaires from the Metadata repository and can link the required questionnaire to the survey and the sample. By adding the message for the survey, selecting the method of the survey (for example, by email), the SAT can begin to extract the questionnaire XML in QPT format from the Metadata repository. If there are any prefilling questions (for example, the name question) in the questionnaire, the SAT will look at the database of the population, and will extract the data (the name of each respondent) from the database and will prefill the questionnaire. Then it will use the QPT html generation tool to generate the HTML visualization page for each individual questionnaire XML file. Then the questionnaire XML and HTML files will be sent to the respondent by email.

Once the respondent receives the questionnaire (XML and HTML file), he can open the HTML file in Internet Explorer with the QPT software installed as plug-in Active X component. The QPT software will link the html to the XML file and will ensure the control of the questionnaire. Once the respondent finishes the filling-in of the questionnaire, he can save the answers back to the XML file. Then the XML file with the data can be sent back to the survey manager.

The SAT will monitor the email server that will receive the returned questionnaires, and will analyse the statistical data of the responses. The SAT can also send reminders to the samples if they have not returned the questionnaire in a given time period.

The DIT tool is used when the respondent wants to extract data from a database to the questionnaire automatically. The respondents need to define the mapping between the variable in the questionnaire and the data located in the database. An SQL statement will be generated for the simple mapping. For

[1] For the respondent, the prefilling question is same as other question except that the answer is already given. The respondent can change the prefilled answer if it is not accurate. However, the questionnaire designer can also define such kind of question as “read-only” in QDT. In this case, the respondent cannot change the answer.

accessing complex data structure in the database, the SQL statement can be manually defined. Once the mapping has been done, the QPT software can connect to the database, extracting the data from the database to the questionnaire when the questionnaire is opened.

4. Common model and mapping of the models

The common model (Figure 4, top picture) is implemented in the metadata repository to enable the storage of the metadata object and the exchange of objects between QDT, SAT and QPT/DIT. Here we have a part of the common model developed with the UML (Unified Modelling Language).

The common model is based on the XML4DR. It is one interpretation of the questionnaire in the data reporting form. The QPT uses XML4DR and the response method is via the web. The XML4DR is designed to be used with web-based forms. In the IQML project, the QPT does not have an object model in UML format. However, the relation between the common model in UML and the object model used in the QPT is very close. This diagram represents the core of the questionnaire in terms of its structure (Set, Domain, Tuple, Data) and its validation (InputControl). Note that this diagram has as its root class the “Questionnaire”. The remainder of the classes have names that are independent of a questionnaire (e.g. there is no “Question” class), as the model represents the structure and related metadata required for data reporting in a general sense: a questionnaire is a specific reporting instrument.

In contrast, because the QDT is focussed on the generic concept of the questionnaire designer, the questionnaire model in the QDT is focussed more on the common understanding of what a questionnaire is. The figure (Figure 4, bottom picture) shows a part of the QDT UML model.

A questionnaire is consistent with the questionnaire body. In the questionnaire body, the possible components are question, section, note and rule. A question is the basic component of the questionnaire. Depending on the question type, it includes one or more variables. Each variable is used to collect a respondent’s answer. Those variables are the same as the variables in the dataset. A note is a piece of text information which can be attached anywhere in the questionnaire. It is often used for the comment, introduction, instruction, footnote, etc. It can be used for the respondent and also for other people involved in the survey. The rule is a navigation rule which will control the flow of the responding questionnaire. A section is a group of questions, notes and rules. The QDT questionnaire model is implemented in the QDT module.

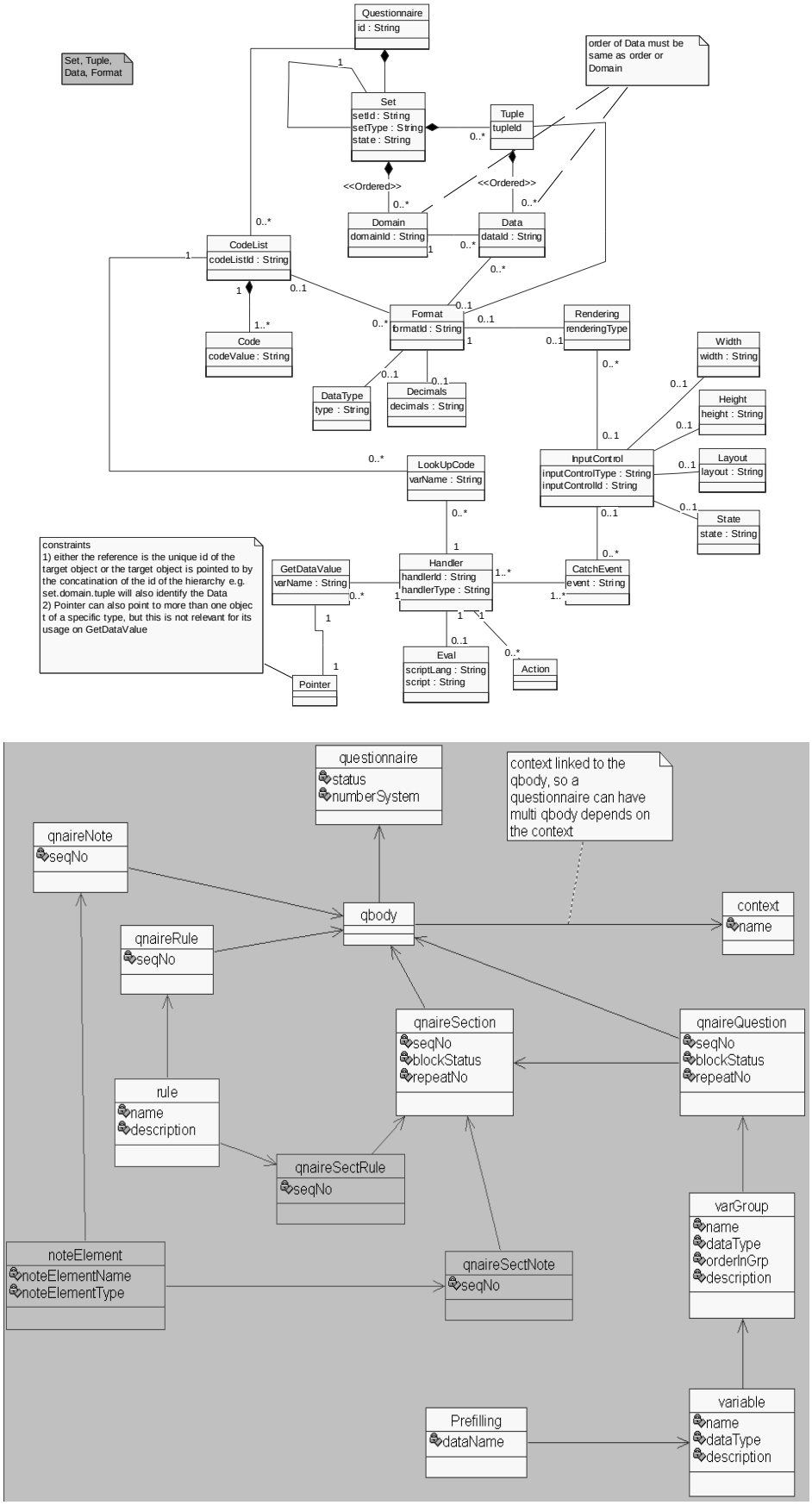


Figure 4. IQML common model (top) and QDT questionnaire model (bottom)

When a questionnaire is designed in the QDT, it will be stored in the local disk as XML format according to the QDT questionnaire model. When the questionnaire is published in the Metadata Repository, it will be converted to the IQML objects according to the IQML model used in the Metadata Repository. The mapping of both models is therefore very important for the interoperability between different software modules. The basic mapping of the objects in both models is listed in the following table. A full mapping has been developed in the IQML project and implemented by exporting the questionnaire from the QDT to the Metadata repository.

Table 1. Mapping QDT model and Common model

Object in QDT questionnaire model	Object in Common model
Questionnaire, questionnaire body	Questionnaire
Section, Question	Set
Response Group	Domain
Variable Group (sub question)	Tuple
Variable	Data
Member of the response group	CodeList, Code

By using this mapping, the questionnaire defined in the QDT is published in the Metadata Repository and is ready for the SAT to use in the Survey.

5. Interoperability

The IQML software suite includes 5 modules developed by 4 partners. The interoperability of the modules is the main issue of the IQML software development. To ensure interoperability, three trials were planned during the life of the project. The trials have tested the software in the real surveys, and then comments and requirements have been collected after each trial. The users' inputs from the real trial experience were very important for the development of each new version of the software.

Conducting the trials in different stages of the IQML project ensured that the software developers considered the interoperability of their software at all times, because each trial required an integrated IQML software system.

As we discussed above, to enable flexibility in the development of each module, each module has developed using its own model. Using the UML to describe the model has helped to define the mapping between different modules; the mapping has ensured that the necessary metadata has been carried from the QDT to the QPT through the Metadata Repository and the SAT.

To ensure the interoperability of the software modules, the common API (Application Programming Interface) was defined in the early stage of the project, and was regularly updated during the lifecycle of the software development. The QDT and the SAT access the Metadata Repository through the common API (see figure 1) to store and retrieve objects. This common API is the key factor for the interoperability of the whole IQML software suite. Two APIs have been developed for the Metadata Repository: Storage API and Repository API (see figure 2). The APIs are implemented with standard Java Servlet technology to ensure the APIs are accessible through the Internet with the HTTP protocol. This enables the QDT, the SAT and the QPT to share the same metadata from the central Metadata Repository through the Internet.

When a questionnaire is exported to the Metadata Repository from the QDT, the objects in the QDT questionnaire format are converted to the objects in the IQML common model format according to the mapping between two models, and then the objects are stored in the Metadata Repository. Later, these questionnaire objects will be extracted from the Metadata Repository to a questionnaire XML file in RDRMES 1.0 format. This XML file will be used in the QPT for the respondent. As different software modules use different models, the mappings between the different models are the key factor to ensure the interoperability of the IQML software suite.

Before each major release of the software, and before the software was used in each trial, much integration work was carried out to ensure the interoperability of the IQML software suite. In advance of the trials, the users received the software for testing, and relayed their findings on the interoperability issues to the software developers so that these problems were solved before the trials began.

The final trial was successfully finished at the end of March 2003. It showed that the IQML software suite is an integrated software system and was used for surveys in three countries.

By defining the module in the UML and defining the mapping between them, we have gained experience of exchanging metadata between different systems. In the frame of the project COSMOS (Cluster Of Systems Of Metadata For Official Statistics, http://www.epros.ed.ac.uk/cosmos), the IQML QDT questionnaire model and the SAT survey model have been mapped to the NESSTAR (http://www.nesstar.com) DDI (Data Documentation Initiative) standard through the COSMOS common core model. It shows that interoperability can also be achieved once the mapping between different models has been fully defined.

6. Standards

To ensure the interoperability and the integration of the software modules, the IQML project has taken advantage of and contributed to the standards making process, both in its own design and in its feedback to the process. We have used the UML as the modelling tool and have used Object languages such as Java for developing the software. We have followed standards developed for interfacing to XML standards, such as XMI (XML Metadata Interchange) and ebXML specifications. We have also standardised the UML model in a recognised standards organisation (e.g. OMG, CEFACT, ISO). Furthermore, we have developed syntax-specific specifications from the UML model to suit the processing or exchange environments (e.g. XML, EJB, CORBA). In addition, coarse-grained objects can be glued together using XML, especially when communicating over the web. More detailed descriptions about the standard effort and contribution from the IQML project can be found in Chris Nelson's ASC paper "The Affect of Standards on Software Component Architecture" [Nelson, C. (2001)].

By using the technologies that conform to these standards, the integration of IQML software modules has proved to be fairly trouble-free and the IQML software suite is now running smoothly despite the five software modules having been developed by four project partners in three different countries.

7. Conclusions

From the IQML project we have gained much experience in producing surveys with new modern technology and in exchanging metadata between different systems. The core model of IQML is the key factor supporting the exchange of metadata between modules and for 'downstream' applications. The IQML project adopted the RDRMES 1.0 standard as the base of its core model, and has also actively contributed to the standards. The QDT and SAT have specific models that reflect their particular functionality. Through mapping their models to the common model, interoperability within the IQML software suites was achieved. The trials in the IQML project showed the success of integration of the IQML software suite. More information on the IQML project and results can be found in the IQML project's website (http://www.epros.ed.ac.uk/iqml).

8. Acknowledgements

Much material for this paper is drawn from the deliverables, presentations, and working documents of the IQML project. IQML material is the copyright of the IQML project and is reproduced with the permission of the IQML project.

Thanks to the CES support team for help in producing this paper, especially Moira Burke and Adam Taylor.

References

Brannen K (2001) Intelligent Use of Metadata in the Questionnaire Design Process. In Proceedings of the NTTSl-ETK Conference, 2001, Crete, Greece.

Nelson C (2001) "The Affect of Standards on Software Component Architecture" in Proceedings of the ASC conference: "The Challenge of the Internet", edited by Andrew Westlake, Chesham, UK, 11-12 May, 2001

Lamb J M "Metadata in survey processing", Proceedings of the EUROSTAT Statistical Meta Information Systems Workshop, Luxembourg, 2-4 February, 1993 ISBN92-826-0478-0.

Taylor A "The QDT: A New Tool For The Future Of Questionnaire Design" accept for the ASC Conference 2003.

About the Author

Defeng Ma is a research fellow in CES, University of Edinburgh. He has worked on several European research projects on multimedia information systems and educational software. His research interests include distributed high performance multimedia information systems, the usage of multimedia products in compulsory schooling, the extended Markup Language standard for intelligent questionnaires and the multi-agent integration of shared statistical information over the (inter)net. He has been involved in various EPROS projects, including IQML, Mission, Cosmos, Codacmos and MetaNet

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

The QDT. A New Tool For The Future Of Questionnaire Design

Adam Taylor & Defeng Ma

Abstract

The notion of conducting an on-line survey remains relatively new, despite the rapid expansion of technological possibilities in recent years. The design of surveys that can be presented and completed on-line presents new possibilities and challenges for the designer. Concepts such as 'flexibility' and 're-use' become important considerations with the possibility that the same questionnaire can be contextualized in different ways, according to the media in which it is presented, or with the notion that the same questionnaire 'components' can be utilised in different questionnaires. The purpose of the software described was to develop a prototype that demonstrated the value of a novel approach to these conceptual issues, leaving open the possibility of further development of the software based on the underpinning conceptual model.

Keywords

Questionnaire design, metadata, intelligent questionnaires, online questionnaires, questionnaire software, survey software

1. Introduction

Background

The Questionnaire Design Tool (QDT) is one module in the software suite that comprises the IQML (Intelligent Questionnaire Markup Language) project. IQML is a 3-year, EU funded project, which finished in April 2003 (IST-1999-10338). The general aim of IQML is to enable the creation and re-use of 'metadata' in the survey process and to innovate in the areas of the design and presentation of surveys.

The QDT was required to interface with the other relevant software modules in the project. These modules were a Survey Administration Tool (SAT), a Repository for metadata, a Database Interrogation Tool (DIT) and a Questionnaire Presentation Tool (QPT). A 'common model' was developed collectively by project partners at the instigation of the project and was refined as necessary throughout the project lifecycle. It identified objects that would be common to the various modules. Where necessary, different models were also developed and used by the modules, according to their specific requirements. A model was developed for the QDT to cater for a questionnaire design package intended to be responsive to the requirements of as wide a range of user needs as possible. A high level view of this model is outlined below.

Building Blocks – a high level introduction to the QDT 'components'

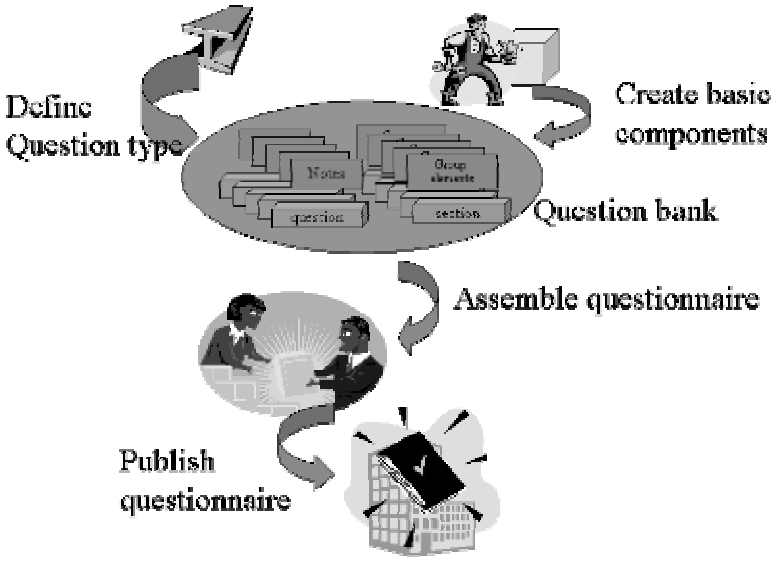


Figure 1. The QDT 'philosophy' for designing a Questionnaire

The conceptual basis of the QDT lies in the notion of creating re-usable 'components' as part of the questionnaire design process. In effect they can be thought of as 'building blocks' where elementary components can be combined and recombined to make more complex ones. Once having created them, the designer will always have these fundamental components of a questionnaire at their disposal.

The basic QDT components are the following:

Question Type – structural 'templates' for questions

'Question Bank' components:

- Element Groups:
  - Response Group – a group of response texts
  - Missing value group – a group of missing value texts
  - Sub-question Group – a group of sub-questions texts
- Note – a piece of text
- Question – a question to be used in questionnaire(s)
- Section – a grouping of questions and / or notes

Questionnaire

'Control Element' components:

- Calculation – a mathematical calculation that is applied to a variable in a questionnaire
- Validation – a validation check that is applied to a variable in a questionnaire
- Rule – a 'navigation' or 'dynamic guidance' rule that is applied to a questionnaire

Context – the context (e.g. sample group) to which a questionnaire is applied:

- Language – the language that is applied to a questionnaire
- Media – the media (e.g. paper, phone, online) in which a questionnaire is published

‘Question Type’ components serve as ‘structure templates’ for Question components by defining the basic characteristics of a Question. These characteristics or ‘properties’ are the ‘data type’ of the response (text, integer, real or boolean), the ‘type’ of response (‘simple’ – typically a text or numerical response typed by the respondent, ‘multiple’ – where the respondent may select multiple responses from a selection or ‘choice’ – where the respondent can select one option from a selection) and whether the Question contains ‘Sub-questions’. The definition of the Question Type properties determines the components that can be included in a Question. Examples of the way in which these properties are used are illustrated in the various questions shown below.

Figure 2.

This question would have a question type where data type = boolean (each response can be checked or unchecked), response = choice (only one response may be selected) and sub-questions = no (no sub-questions are used).

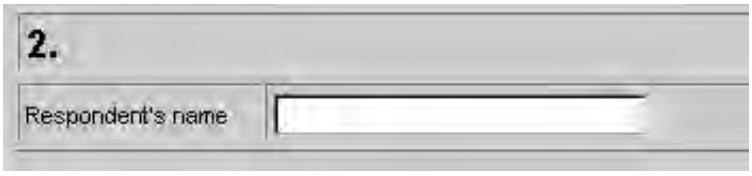


Figure 3.

This question would have a question type where data type = text (text response required), response = simple (all ‘typed responses are ‘simple’) and sub-questions = no (no sub-questions are used).

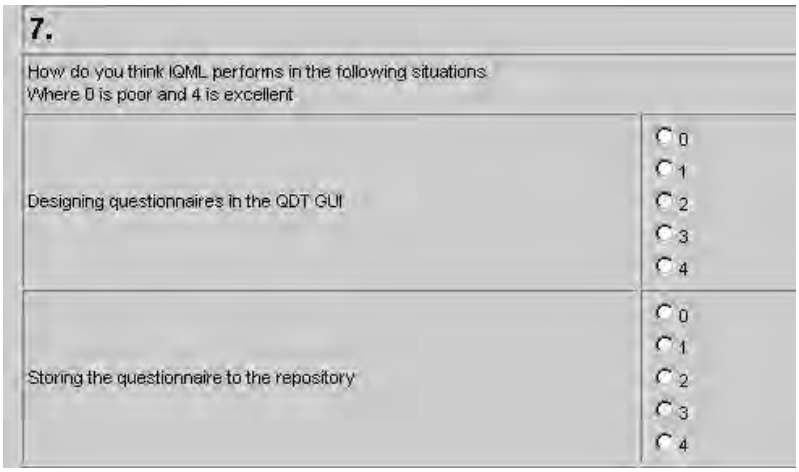


Figure 4.

This question would have a question type where data type = boolean (each response can be checked or unchecked), response = choice (only one response may be selected) and sub-questions = yes (sub-questions are used). The Question uses both a Sub-question group – in the left column - and a Response group – in the right column

The 'Question Bank'[1] contains Questions and associated components. 'Element Groups' are collections of text elements – most commonly they are grouped into 'Response' or 'Sub-question' groups (see Fig. 4, above). 'Notes' – essentially pieces of text - may form part of the composition of a Question (an example of a re-used Note component ('Where 0 is poor…') can be seen in Figures 4 and 5). 'Sections' are collections of Questions grouped together. Along with Questions, Sections form the basis of Questionnaires.

'Questionnaires' are collections of Questions, Sections and Notes that can be arranged in the desired sequence by the designer. By connecting with other modules in the IQML suite (namely the Repository and the SAT) the Questionnaire can also be populated with 'pre-fill' information which is taken from the sub-sample that is to be used in the survey. The Questionnaire can also have 'Control Elements' (see below) applied to it.

'Control Elements' are, as suggested by their name, 'controls' that the designer may introduce to the Questionnaire. These can be 'Calculations' - arithmetical functions performed on input values, 'Validations' – checks on input values, and 'Rules' – the ability to specify 'dynamic guidance' (or 'go to' instructions) within the Questionnaire, dependent on the input values provided by the respondent. All Control Elements are similar in that they are defined as free logical / mathematical expressions, with variable values from the Questionnaire subsequently 'applied' to their parameters.

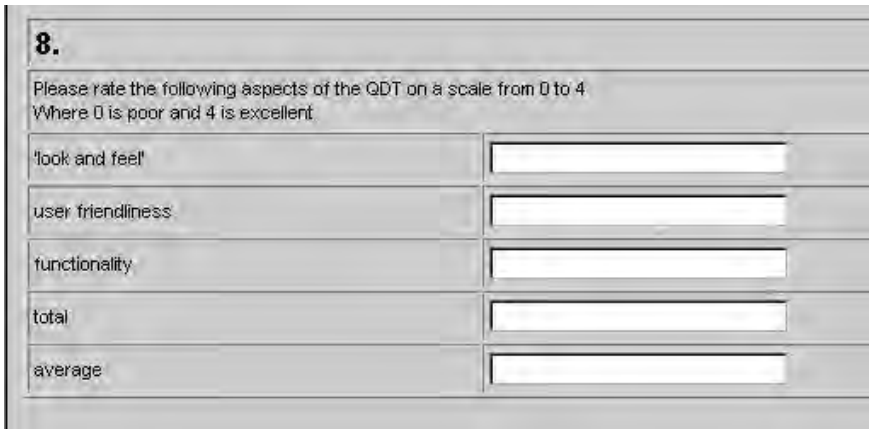


Figure 5.

Calculations could be applied to the 'total' and 'average' fields in this Question. Validations could be applied to ensure that the values entered are within the desired range, or to compare the input value against other values.

The methodology outlined above created an approach to the process of questionnaire design whereby the designer would be forced to think about the design of their questionnaire from the 'bottom-up'. First they create the Question Type 'templates', then any other components necessary for the creation of the desired Questions, then the Questions themselves and so on. What becomes evident through this approach is that many components that may have been created with a specific Question in mind will subsequently be re-used many times creating other Questions.

While in practice the IQML project has focussed on outputting questionnaires in digital format, the QDT model allows for the output of questionnaires in a variety of media formats (e.g. paper, web-

[1] The concept of re-usable 'Question Banks' exits elsewhere in the research community – for an example the CETIS (Centre for Educational Technology Interoperability Standards) site at http://www.cetis.ac.uk is building specialist question banks for a variety of subject areas.

based, telephone etc.). This flexibility would allow the questionnaire designer to design their questionnaire for a specific media format, or alternatively, to output the same questionnaire in different media formats, with the further possibility that different 'contexts' (e.g. the wording of specific questions) may be applied to the same questionnaire when output in different media formats.

QDT uses flexible, extensible technologies

The QDT software has been designed and implemented using the Java programming language and the Extensible Mark-up Language (XML). Using these technologies was strategic – the reasons for which are outlined below.

The advantage of using Java lies in its 'platform independence'. Programs written in Java function by using the Java 'virtual machine' – which may be installed on just about any computer. Hence 'platform independence' means that software written in this language will not be bound to any particular computing environment – for example an exclusively PC or Macintosh environment – as might have been the case had another programming language been used. Java also has the advantage of offering standard Unicode support, to enable the support of non-English language and characters, like Norwegian and Greek – a highly relevant consideration given the identity of the project trialists (see section below).

XML technology enables the development of documents capable of holding structured information, such as that produced by the QDT, but allows for an arbitrary structure – as demanded by the QDT model. The 'extensibility' of XML also ensures that the QDT program and it's underlying model can grow and incorporate new and developing demands that may be made of it, without having to be re-designed from the 'bottom up' – an important consideration in an environment where new requirements of the model may emerge and need to be incorporated. XML documents are also ideal for transmission via the web and as such were a very convenient mechanism for the transferral of data between the IQML modules as outlined above.

Through the experience of the IQML project, it has been demonstrated that the QDT model can be mapped to software supporting the distribution of online questionnaires (in IQML, this was the SAT). Compatibility with this method of survey distribution and collection is important with the probable expansion of this means of distribution / collection in the future.

2. Implementation Experience From the IQML Project

The project trialists

The QDT has been designed with the intention that a survey producer from any organisation may use it. Due to the fact that the IQML project trialists and partners are drawn from a variety of survey producing backgrounds, the QDT has been designed to cater for the generic needs of the survey producer.

The project trialists for the IQML project comprised two National Statistics Institutes and one academic institution. They were the Central Statistics Office (CSO) of Ireland, Statistisk Sentralbyraa (SSB) of Norway and the National Technical University of Athens (NTUA), who assisted the National Statistics Office in Greece with the implementation. On the development side, the developers of the SAT, DESAN Research Solutions, also specialise in the design of market research surveys for non-commercial organisations, and as such were also able to lend their expertise to the project. These

organisations provided considerable experience of questionnaire design and of producing a wide range of survey types.

The input from each organisation at the requirements gathering phase of the project was vital in identifying the functionality required from the tool. Through careful analysis of sample questionnaires provided by the trialists, and discussion of required functionalities with each organisation, it was possible to gradually develop a picture of the software tool they wanted.

The project conducted four trials of the software, with functionality added incrementally with each release. This method of 'prototyping' proved to be effective in getting quick user feedback for various functions, but principally for identifying missing functionality and for generally stimulating discussion between the users and the developers.

As mentioned above, in the course of their everyday work, the trialist organisations produce a wide range of survey types. For the final IQML trial CSO used the modules in the suite to design, produce, distribute and collect a Quarterly Accounts Inquiry to Industry, SSB conducted an internal survey on employees (lack of resources prevented them from conducting their planned survey of municipalities) and NTUA conducted a direct survey on small enterprises.

What was actually implemented and why?

In keeping with the traditional pattern of software development, not all of the originally planned functionality was implemented in the final version of the QDT software. Due to lack of resources, impending deadlines and the absence of complimentary functionality in some associated modules, decisions had to be made prioritising certain, indispensable functionalities over those that could be considered peripheral or which were not going to have the necessary supporting functionality in other modules in any case.

So what was considered to be indispensable and why? The primary aim of the IQML project was to demonstrate that the prototype software modules described above could inter-operate as an integrated software suite in a realistic trial scenario. On the basis of the questionnaires submitted for the final trial it was decided between the project developers that such functionality as to satisfy the demands of these 'final trial' questionnaires should be implemented. As the final trial (and indeed all of the trials) had concentrated on the visualisation of the produced questionnaire in an on-line format via the QPT software, the functions that were given priority for inclusion in the final trial were inevitably those that supported this aim. One consequence of this was that functions that would have given the designer control and choices over the media format of the questionnaire, while fully developed in terms of the QDT model, were jettisoned in the implementation. Given that, for the purposes of the trial, the output of the questionnaire was only ever going to be realised in one format (i.e. online), the resources required to enable output in other formats could not be justified. Likewise, the final trial questionnaires were only ever going to have one 'context', so although this functionality could be extremely useful for the production of certain surveys in other circumstances, it was never going to be used for these ones.

Another factor limiting the functionality that could be included was what would be supported in other relevant software modules – in the case of the trial scenario this was the tool through which the questionnaire was visualised, the QPT. The most significant functions to be lost in this way from the final implementation were the capability to set text parameters in questionnaire questions based on responses given by respondents to previous questions and the automatic setting of response values based on the values given by the respondent to other questions. Both of these omitted functions were

included in the final trial questionnaires but could not be supported by the QPT. Given that these limitations were known, it was not considered sensible to use the resources that would have been required to implement the functionality in the QDT. Despite these limitations to the functionality of the prototype developed for the final trial, the absence of the aforementioned functions did not represent a 'trial stopper' scenario for the trialists and so the trials proceeded with the questionnaires slightly modified to accommodate the missing functionality.

Trial Results

In the trial report "the aim of the evaluation of QDT was to get as much feedback as possible on usability, layout of screens, concepts and terms used." While all of these considerations are obviously highly important in the evaluation of any software product, given that the software being evaluated here was never considered to be anything other than a prototype (as opposed to a professional release standard), the emphasis here will be on the evaluation of the conceptual aspects of the program, as opposed to the finer points of it's 'usability'.

Once the QDT package was installed and the users were able to commence the 'conceptual' work of the trial, the objective was to translate a questionnaire that had been produced 'on paper' by questionnaire designers from the trialist organisations, into the QDT. From the outset this would involve a transfer of the concepts implicitly expressed in the paper version of the questionnaire into the conceptual framework of the QDT. The most fundamental task in the creation of a questionnaire could be considered to be the creation of a question. That this was the approach of at least some of the trialists is clear from the statement that

"All system users found it quite easy to *create a question* and provide all the required information, although the *modifications of questions* was a more difficult task. When modifying an existing question, it is not possible to change the *question type* or any associated *element groups*."

This comment demonstrates that the system users, at least to some degree, grasped the conceptual framework of the QDT, namely that before constructing a Question component it is first necessary to create the Question Type 'template' for the Question and also to create any Element Groups associated with the Question. It also reveals some of the biggest limitations of the system - that structural modification of questions was not possible. To change the Question Type of a Question is, in terms of the QDT model and the programming involved, problematic, but not impossible. In terms of the QDT 'philosophy' it is undesirable for the reason that it would be better for the designer to begin with a clear understanding of the structure of the Question they are creating. The Question Type is the 'foundation' on which all other components that need to be included in the structure of the Question are based, so to comprehend what this is at an early stage is obviously the best way to use the program. This said, however, in the real world questionnaire designers are human and inevitably make mistakes, so the ability to change Question Type would be an enhancement to the program.

The comment concerning the Element Groups reveals another limitation of the program in its present state of development. Changing a 'peripheral' component like an Element Group should, in principle, be perfectly possible, as it does not change the structure of the Question (one can easily imagine a scenario where, for example, a questionnaire designer might wish to change the response categories for a question from "Yes / No" to "Yes / No / Don't know").

Other limitations of the QDT / IQML package were also mentioned in the report. Notable for the QDT were comments related to the easy visualisation of the Questionnaires created by the QDT. This would certainly be a welcome enhancement of the package, as users would naturally prefer to have an easy

visual aid when constructing their questionnaires. The ability to 'see' what you are producing can be an obvious advantage. To have such functionality would have required that the QPT be automatically invoked by the QDT, however this was not possible as the two software modules were not dynamically linked in such a way due to the resources that this would have involved. Instead it was necessary to export the QDT metadata to the Metadata Repository, converting it to the model used by the QPT, from where the QPT could pick up and display a converted XML script.

Another, related, limitation was the lack of formatting features for the Questionnaire, via the QDT. Theoretically such features should have been possible but due to lack of resources and the prioritisation of other features such an enhancement was not possible.

In the most general terms the evaluation of the QDT has been positive. While several comments in the evaluation relate to the desirability of certain features that could enhance the 'user friendliness' of QDT, these concerns, while valid, are not of primary importance. Overriding these issues, which may be viewed as essentially 'cosmetic' in the sense that they do not affect the basic functionality of the program, is the bigger issue of whether the users of the system gained a *conceptual* grasp of what the QDT was attempting to achieve. Based on the evidence gleaned so far from evaluations of trial scenarios it would seem that the system users could see the value of this approach. As the evaluation states: "Overall the whole environment of the package was considered to be *user friendly* and the general feeling was that in the future the system would be extremely useful, if it is enhanced with additional features and more functions."

3. Future Possibilities

Extra Functionality

Based on the concepts already included in the QDT model, there are many possibilities for the future development of the QDT, although any further development would depend on new funding.

The first, and perhaps most obvious, area in which the QDT could be further developed would be in implementing the 'online' functionalities that were not fully implemented in the IQML project, due to the lack of the required functionality in the associated IQML module, the QPT. 'Online' functionalities should be taken to mean those functions that would only be of relevance to an online questionnaire.

Some of these functionalities already exist in the QDT interface that was developed for the final trial, namely:

- the ability to specify 'navigation' or 'dynamic guidance' rules was supposed to be an available functionality for the final IQML trial
- the ability to specify which questions within the questionnaire should be 'blocked' to the respondent, depending on navigation rules

Foremost amongst functions that have been modelled by the QDT, but not implemented, are the ability to:

- define rules that could automatically set data values in one field, dependent on the value that the respondent has entered in another field
- take text 'parameters' from responses to questions and include them in the text of other questions in the questionnaire

- set different 'contexts' for a questionnaire – in practice this would mean that the same questionnaire could have, for example, different texts for the same question, that could be used depending on which 'context' the questionnaire is being completed (e.g. by interview or by self-completion)
- related to the 'context' function described above, 'language' and 'media' functions have been modelled to allow for the possibility that the same questionnaire may be produced in different languages (a pertinent functionality for many National Statistical Institutes) or presented on different media.

As with those functionalities cited which have been implemented in QDT, but not the full IQML suite, those described above could be relatively easily incorporated into the QDT however the main issue would be to find complementary modules with which this fully functional module could inter-operate.

Interfacing with other modules

The experience from the IQML project has demonstrated that the QDT model can 'map' or 'interface' with a model underpinning an associated software module. Establishing that this is possible is important when considering the possibility of developing the QDT functionality and 'inter-operating' with other software modules. That some of the richness of the QDT model could not be incorporated into associated models in the IQML project is not a concern as this was due to the limitations of the other models – where there was conceptual 'convergence' between the models, the process of mapping was found to be straightforward.

The QDT is also being utilised as part of the COSMOS 'cluster' project (IST-2000-26050) (http://www.epros.ed.ac.uk/cosmos/index.html), where XML output from the QDT is mapped and converted to the Data Documentation Initiative (DDI) standard (http://www.icpsr.umich.edu/DDI/index.html), from where it is utilised by the FASTER (IST-1999-11791) project. This serves as further evidence that the QDT can interface effectively with other relevant software modules in the statistical domain.

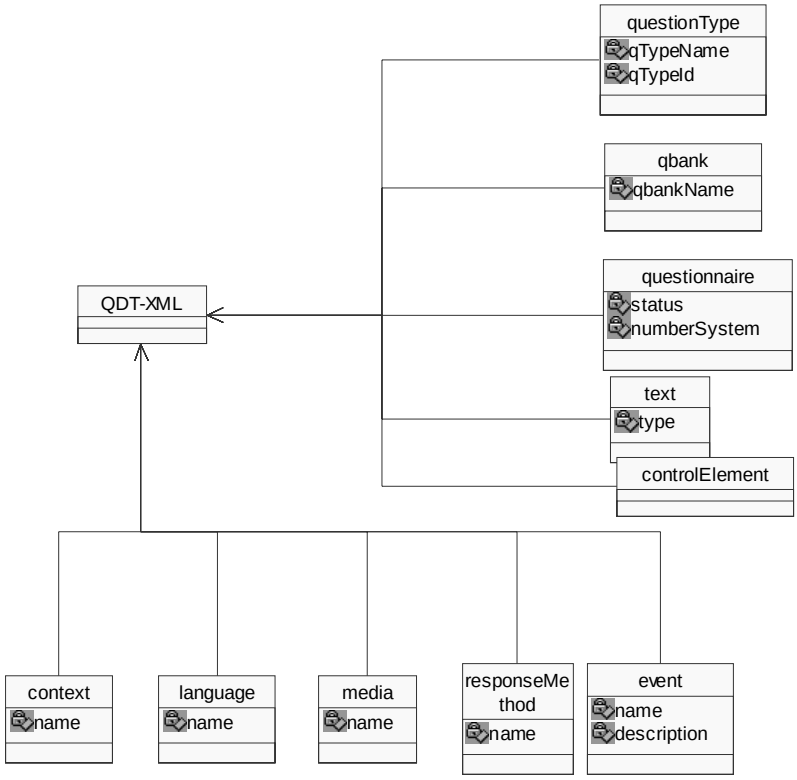
4. Concluding Remarks

As will be obvious from the previous discussion, as a piece of software, at this moment, the QDT is far from complete. Whether or not it will reach 'completion', in the sense that all of the potential functionalities described are actually implemented and ultimately utilised by the types of users described, the conceptual insight that has been gained from the development of the QDT model and the experience of developing the software through the lifecycle of the IQML project, has value. The main 'purpose' of the QDT has been to enable the creation of questionnaires and associated metadata using a different conceptual approach to the orthodoxy. The preliminary results available suggest that, after a period of 'conceptual acclimatisation', users of the system have come to understand the philosophy that underpins the architecture of the system and to use it accordingly – namely to design questionnaires from the 'bottom-up' using the 'component' based approach described here. Whether it can be developed beyond the 'prototype' scenario remains to be seen, but nonetheless it is to be hoped that the development of the QDT has advanced this area of knowledge.

5. Acknowledgements

I would like to acknowledge the work of Joanne Lamb. Much of her work on the IQML project has been drawn on in the production of this paper.

Appendix - QDT UML Model



This is a 'high level' view of the QDT model. More detailed representations of the model may be obtained on request from the author.

References

Evaluation Report of Final Trial, IQML project report, June 2003 (http://www.epros.ed.ac.uk/iqml/deliverables/deliverables.html)

http://www.icpsr.umich.edu/DDI/index.html

Ma D 'Developing a Software Suite for Intelligent Questionnaires: Interoperability and Standards', ASC Conference, Warwick, U.K., 17-19th Sept., 2003

About the Author

Adam Taylor is a Research Fellow at the Centre for Educational Sociology (CES), which is part of the University of Edinburgh. He has worked on the IQML and COSMOS projects, contributing to the design and implementation of software for those projects.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*.

Batch Tabulation. A Premature Obituary

Keith Hughes

Abstract

The paper traces the history of Batch Tabulation through some major landmark programs, reflects on the current use of Batch Tabulation software, and makes some cautious predictions about its use in the future.

Keywords

Batch Tabulation, AGSP, Merlin, Quantum, Snap.

1. Introduction

When Market Research and the Computing Industry came together in the mid 1960s, their first child was Batch Tabulation software, a fast and reliable replacement for counter-sorters and electronic adding machines in the production of sometimes vast volumes of cross-tabulations. Nowadays, computers are used in every part of the Research process from designing questionnaires through to publishing reports on the Internet, and by far the greatest growth has come in the area of data collection. Looking at the ASC's software catalogue, or browsing Tim Macer's web site[1], you might conclude that the chief purpose of Market Research today was to fill data warehouses with information that someone might want to look at later. And listening to the leading Market Research software supplier, you might infer that in the future the researcher will carry out any odd bits of data analysis that might still be required by eccentric clients, and that the professional survey analyst is now at least an endangered species—although it's not yet clear that any reputable zoo or wildlife park is ready to attempt a breeding programme. Having been involved in Batch Tabulation for more than 35 years, I feel I am as qualified as anyone to speak its epitaph, but I hope that this essentially historical paper may suggest that its day is far from over.

2. AGSP and Merlin—Survey Analysis for Programmers

The history of Batch Tabulation in the UK really starts with AGSP (the Atlas General Survey Program), designed by a team led by John Williamson at ICT in 1965. AGSP was a high-level programming language; it was used only at ICT (later ICL) and ran only on Atlas computers, of which there were only ever three built. In its heyday, the AGSP team consisted of John, two

[1] http://www.meaning.uk.com

development/support programmers, and five users. AGSP was what was then called a "high-level" (i.e. as opposed to machine code) language, and in addition to standard high-level language components such as variable declarations and assignments, and flow-control statements, AGSP had specific "table definition" statements and, crucially, its own macro language, making possible the specification of hundreds of tables in a few simple statements. It took a novice several months to become fully productive in AGSP; there was no formal training—"Read the manual" was all we were told—and new programmers were put straight on to "live" jobs. (ICL lost a number of clients that way!)

AGSP had its competitors—IBM's SAP program and Pritchard, Brown & Taylor's "Sable" were probably the major names—but in terms of power and features, AGSP was streets ahead of them all. ICL classed the AGSP team as "programmers", and vacancies usually arose as a result of staff being transferred to other programming teams within ICL.

AGSP passed out of use when the last Atlas was de-commissioned in 1973, and the same year saw the first appearance of the program which John Williamson was later to describe as "AGSP's natural successor"—Merlin. As one of Merlin's original authors, I can safely say that most of the user interface was lifted, with slight improvements I hope, from AGSP: the emphasis was strongly on power and features, and it was intended to be used by people with a programming background or at least programming aptitude.

Merlin is not only still around today, but is still under development, and new releases, incorporating major new features, appear every year. Sales (I'm pleased to be able to tell you) were actually up last year. It's widely used, not only in the UK but in mainland Europe, America, and the Pacific Rim. In September 1995 it won an ASC "shoot out" as the most powerful package on show. It's used in analysing passenger surveys for the Tyne & Wear Public Transport Authority, freight shipment surveys for DHL, and Royal Mail and British Telecom customer surveys: wherever data structure and analysis requirements are of a high order of complexity, there's often a Merlin programmer at hand.

Merlin continues to be the main representative of the tradition started by AGSP—survey analysis for programmers. Although the term "specwriter" (of which more in a moment) is now more commonly used as the Merlin user's job description, writing Merlin shares most of the essential elements of mainstream programming, notably the fact that the user has considerable scope to choose the most efficient way of writing his script—efficient both in the sense of machine use and of time taken in script-writing.

3. Quantum—Survey Analysis for Specwriters

The mid 1970s saw the emergence as market leader of what has become the best-known and most widely-used Batch Tabulation program, Quantum. Designed and mostly written by Ed Ross, it was the first such program ever to be sold for use by an end-user—the Post Office, as it happened—and therein lies the story of its success. Relative to AGSP or Merlin, it is exceedingly simple to use and it can be picked up very easily. Although it's driven by a script, the Quantum language is declarative rather than procedural (however, there is a procedural section if you need to recode your data). The rules governing the definition of tables are straightforward if arbitrary (e.g. "N10" denotes a base row or column), and there is often only one way to perform a particular task.

Quantum made "in house" survey analysis practical, and gave rise to a whole new profession. I think it was Jeremy Pritchard, of Pritchard, Brown & Taylor, who first coined the term "specwriter" in the late

1960s to describe someone who analysed surveys for a living; it's worth remarking here that although some practitioners regard this classification as a management manoeuvre to pay them less than programmers, clearly a different skill set is needed for this particular job, and not all programmers make good specwriters, and vice-versa. The emergence of Quantum defined that skill set, and went on to boost the job advertisement section of Research Magazine by several pages.

Another important innovation was the choice of C as the Quantum source code language. This meant that the program was portable to a wide range of computers and platforms, significantly opening up a considerable market for the product in the days before we learned that Windows Is The Only Operating System. During the late 1970s, Quantum's chief competitor was PTT's "StarTab", which was more in the AGSP tradition (i.e. more of a programmer's tool), but this wasn't the only reason for their taking the No. 2 spot behind Quantum: StarTab ran only on Digital computers (PDP and VAX). Quantum ran on VAX, Prime, IBM, Honeywell and indeed almost anywhere that C was supported. And when the 1980s came along, Quantum was ready for both the Motorola-based and the Intel-based Unix/Xenix micro-computers.

In 1997, Ed Ross sold Quantum to SPSS, and product development has slowed as announcements about a brand new system to replace it and their other market research software products, Dimensions, have multiplied. Although it's not yet clear whether Batch Tabulation will be one of Dimensions' many offerings, SPSS say that they will continue to recognise the need for this kind of facility, and it will be interesting to see how they meet it. Given the huge number of Quantum users world-wide, clearly any new Batch Tabulation product from SPSS will be met with great interest, and a lot of experimentation; but their bigger concern must be how, in what seems to be a time of significantly reduced IT investment, they will persuade customers to switch from a system with a proven track record and a huge human resource base.

4. Snap—Survey Analysis for Everyone

The dawn of the micro-computer age in Batch Tabulation in the UK can be dated to March 1980, when Peter Wills and Stephen Jenkins demonstrated the first version of Snap on a Tandy Model 1 at the MRS Conference. Versions for CP/M and MS-DOS appeared in the years following, and the program eventually reached Windows in 1992. The big change from previous systems was that users could specify their tables interactively—initially as a question-and-answer dialogue, and then when GUIs appeared, as a form-filling exercise. No scripting, no syntax—so no need for extensive training or computer-oriented experience.

Another important Snap innovation was the use of inverted (strictly speaking "transposed") data files; initially this was done to overcome speed problems on the early micro-computers, but retention of the feature means that Snap is very fast on today's PCs. (In fairness, it should be said that Quantum's "flip" files first appeared at around the same time, apparently in one of those happy co-incidences that plague the researches of historians.)

Snap was a major breakthrough, and as soon as Windows arrived, vendors of scripting systems quickly followed suit with similar GUI systems (e.g. Quantime with Quanvert, Merlinco with Merlinplus). Typically these systems draw heavily on the approach of the "parent" program—a graphical user interface to the parent making use of a subset of parent facilities, with the underlying disadvantage that the user has to have some basic understanding of the parent structure and conceptual

approach. Snap had the advantage of being able to take a fresh approach, unhindered by legacy considerations.

There are significant drawbacks to the GUI approach. Anyone who has used both a scripting and a GUI system to set up tables will tell you that the scripting method is quicker. This has a lot to do with using both a mouse and a keyboard for GUI. There's also an analogy to be drawn here with sketching freehand versus painting-by-numbers, but I think I will leave the reader to ponder it!

Another drawback is the level of sophistication that can be achieved in any system that does not use scripts: there is a limit to the number of menus, modal forms etc that can be displayed without leaving the user lost in the forest, with no clear idea of where he is or which way he's headed. What starts out as an obviously simpler, clearer approach can quickly become an untidy tangle when features proliferate. (I'm not suggesting that Snap is guilty of this, by the way: their sensible policy is to restrict the features and preserve the simplicity.)

But to draw attention to these disadvantages is to miss the point of Snap. It opened up the production of survey tabulations to almost anyone who could read one: producing table reports no longer had to be outsourced to a specialist, be he specwriter or programmer. The result is that Snap now has so many users they've given up holding User Group meetings, as Peter Wills is reportedly unhappy with the acoustics at the only possible venue left to them, the Royal Albert Hall. But, interestingly, Snap has made hardly any inroads into the user base for Merlin or Quantum; it seems that either an organisation feels it doesn't need the power of Merlin or Quantum for its analysis needs, or it wants Snap for end users in addition to Merlin or Quantum for its table production team.

5. Batch Tabulation as it is today

The products reviewed in this paper have been selected purely for their historical significance, and I wouldn't want to give the impression that Merlin, Quantum and Snap are all that's around today. Programs with their roots in the 70s, such as StarTab, CfMC's Mentor, and QPSmr are all still in active use, as are some of the early US micro packages such as Uncle and Microtab. And outside of the commercial market research world, SPSS Tables is still one of the biggest players.

New players in the survey software game during the last dozen or so years have tended to major on data collection software, not surprisingly given the fact that there have always been more people involved in collecting data than in analysing it, and more people means more licences means more money. Some vendors (e.g. Converso) offer a very limited analysis facility, and recommend a partner company for analysis software. Others such as Askia and Voxco offer more sophisticated facilities and very attractive-looking presentation, with an underlying scripting language that can be used for complex recodes, but typically the scripting language is not used to control the tabulation process itself, so flexibility of presentation and content can be fairly restricted. Features such as netting rows below minimum levels, showing four types of percentage per cell, and user-controlled presentation of significance testing tend to be found only in the older programs, where they have appeared largely in response to user demand.

But the demand for batch tabulation seems hardly to have diminished. Commenting on a recent study conducted among AURA (Research Users' organisation) members, AURA Chairman Leslie Sopp told me: "I think it is fair to say that nearly one in five client research departments still only take their data in "book" form, but that as many now only take their data electronically, and many of those who take

their data in “Book” form will want to run their own further analysis. The majority do both, and therefore the days of tabular output from agency to client are not past, but are definitely declining”.

Informal research amongst Merlinco’s research organisation clients indicated that most still regard the book of tabs to be the main deliverable on research studies. And traditional tab bureaux such as IDA and ATP still derive the bulk of their income from bulk tabs.

It’s certainly true that a significant and growing number of end users require data as well as tables, in a format suitable for handling by one of the end user interactive packages such as Quanvert. These programs are a boon to specwriters as well as researchers, because extra tables on a project that was completed weeks ago can be a major irritant in a specwriter’s life, and far from seeing interactive table programs as a threat to their livelihood, most specwriters are glad to be free of the drudgery of relocating all the last-minute edit, recode and weighting files they used on the main tabs run.

What specwriters do see as a threat to their livelihood is the growing tendency of research companies to out-source their tabulation work overseas. A number of tab agencies have sprung up, mainly in India and mainly using Quantum, which seem able to produce bulk tabulations at a considerably lower price than the researcher’s in-house DP department or their usual UK supplier. Discussion of that topic is outside the scope of this paper, but it’s relevant to notice here that these companies have come into being in the last few years on the premise that batch tabulation is a viable enterprise on which to build a new company

6. Conclusion—The Future

SPSS are still far and away the leading suppliers of Batch Tabulation software, so I asked Richard Kottler of SPSS for his views on its future. Like Leslie Sopp, Richard takes the view that demand is declining, but very slowly, adding that he has no illusions that it will die in his lifetime. (As a less leading supplier myself, I wish him a long life!) He sees suppliers and users of research data as gradually becoming more sophisticated, and he draws attention to the change in the pitch from researchers to end-users as reflected in adverts in Research Magazine, where the trend is towards offering services using words like "insight", "perception", "value add", "interpretation", and where the deliverable is more consulting oriented and less of a packaged product. Whilst agreeing that the style is changing, I feel less sure about the substance, and I question how a researcher is realistically going to be able to offer those things, or how an end-user is going to be able to mine down to the level in his data where those things become possible, without a hard core of “door-stopper” tables to act as a basis for this kind of illumination.

And so, finally, to the tools themselves. In a recent article, Tim Macer (2003) stated his amazement that people were still using the analysis tools that were current when he entered the business, and I would certainly agree there have been no significant developments in the field since Snap—although we must remember that Dimensions’ contribution in this area is not sufficiently developed for comment here. But I think there may be other reasons for this apparent lack of development. Although software companies may wish it were otherwise, what users really really want—more than new products with flashy GUIs and a comprehensive renaming of all the features they thought they had just about got to grips with—is robustness and stability. Since Microsoft produced a version of Windows that didn’t need rebooting more than once or twice a day, sales of new versions have dropped off dramatically—sales of XP are well below forecasts. (Hence their recent shift in focus from sales to product licensing revenue.) The products reviewed here are certainly stable, and even “mature”, but

that doesn't mean—as some software vendors imply by the term—obsolete. They are very much in use today, they cover effectively all the kinds of people who might want to use such tools, and all kinds of projects. I believe that the focus amongst Batch Tabulation product vendors has to move on from generating new products for old tasks to attending to the level of development and service that can be offered for existing products. We need to learn from Microsoft that what the customers want is not new ways of tabbing Question 1 by Question 2, but the same way only faster, with more short cuts, fewer bugs, more help for the user and above all, Connectivity (through triple-s XML, of course) to ensure that the right tools can be applied to the right tasks quickly and painlessly.

References

Macer, T (2002) "Sound the alarm. The world has moved on but MR tools have not." *Research magazine*, January, issue 440.

Westlake, A (2003) Ed, ASC Software Register, Association for Survey Computing, ISSN 1475-8121

References Acknowledgements

The author would to express his gratitude to Ed Ross and Steve Jenkins for reading and criticising the passages on Quantum and Snap, and to Leslie Sopp and Richard Kottler for allowing him to quote their views on the current state of and future prospects for Batch Tabulation.

About the Author

Keith Hughes is Development Director of Merlinco Ltd. He has worked in survey software since 1967, and in 1973, he co-authored the first version of the tabulation system that has since become MERLIN. He formed Merlinco Ltd in 1991, and in the same year became a founding member of the triple-s committee. Previous papers for the ASC include "*Twenty five years of dirty data*" (Bristol 1991) and "*triple-s XML: A standard within a standard*" (Edinburgh 1999).

He may be contacted at keith@merlinco.co.uk .

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Using Mainstream Standards for Survey Tabulation and Reporting

John Lyon

Abstract

Most delegates will be familiar with the many proprietary tabulation systems available. When we set out to build Vector, a web-based cross tabulation system, we broke the problem down into the two classic phases, aggregation and manipulation. However, we decided that, rather than developing yet another proprietary set of code for table manipulation, we would make use of a mainstream standard, MDX (Multi-Dimensional eXpressions), which is utilized in OLAP within Microsoft Analysis Services (part of SQL Server), as well as by other major vendors. While OLAP may not seem immediately obvious in application to survey analysis, it permits the manipulation of aggregated data, which is a necessary phase in all but the most simple of cross-tabulation requirements.

This paper examines the basic structure and functionality of MDX and the applicability of the language to survey analysis.

Keywords

Survey tabulation, MDX, OLAP, XMLA

1. Introduction

In the past 12 years I have worked on a wide range of systems for delivering market research data to end users. If I had to identify a single common theme it would be using mainstream (non market research) software to build interactive systems for reporting market research data. The fundamental problem is trying to find a flexible way of storing and manipulating market research data that is accessible from modern development tools but retains the efficiency and reliability of existing production processes.

This paper presents some thoughts on how one relatively new part of mainstream database technology, the MDX query language, could offer some interesting solutions to this basic market research reporting problem. This paper examines the theory of applying MDX, but I must stress that I have found it very valuable in practical software development: it has been used as part of ATP's Vector web tabulation software, as well as in a system to generate intranet page sets from pre-calculated tables.

2. Back to Basics

I think it goes without saying that the most efficient and reliable way to produce a reliable set of aggregated market research data is to use one of the traditional batch processing crosstab. These

systems are mature solutions to a well understood problem and the processes for specifying and checking the output have become basic skills for almost everyone involved in market research.

However, using these systems within highly automated or interactive reporting solutions does present some problems. They do not produce output with enough meta-data to be easily understood by other applications and they do not expose their functionality through APIs. They are highly efficient "black boxes" for transforming raw data into printed tables but they were not designed to allow developers to peek inside and take advantage of the engines that build the crosstab.

3. Building a crosstab

Across the whole range of reporting techniques from script driven crosstab batch processing to building Excel charts from pivot tables, the process can always be thought of in terms of three distinct operations—aggregation, table manipulation, and presentation. I would define these as follows:

- Aggregation—any operation that requires a pass through the raw case data.
- Table Manipulation—any operation that does not require re-aggregation and is not specific to any presentation format.
- Presentation—any operation that is specific to the way the data is being displayed in a particular application.

Example 1

The following simple table is run against the Museums SPSSMR demo quanvert database. It is an unweighted table of respondent age by gender with column percentages.

		Total									
			Age of respondent								
		Total	Total	11-16 years	17-20 years	21-24 years	25-34 years	35-44 years	45-54 years	55-64 years	65+ years
Total		602	602	38	82	95	192	91	55	33	16
	Gender of respondent	602	602	38	82	95	192	91	55	33	16
	Male	339	339	23	50	51	108	49	32	16	10
		56%	*56%*	*61%*	*61%*	*54%*	*56%*	*54%*	*58%*	*48%*	*62%*
	Female	263	263	15	32	44	84	42	23	17	6
		44%	*44%*	*39%*	*39%*	*46%*	*44%*	*46%*	*42%*	*52%*	*38%*

Figure 1 A typical simple cross-tabulation.

The stages required to produce this table as it appears here in this document are:

- Aggregation—produce the table of unweighted counts
- Table manipulation—add in the column percentages
- Presentation—generate a table in MS Word and displays the results

This is the most trivial example possible showing the most common market research table manipulation, adding percentages. Obviously, any crosstab system would have no problem creating this table, but, in the real world of market research reporting, things are usually a bit more involved.

Example 2

A chain of cheese shops across the UK run a survey measuring satisfaction across twenty measures of customer service on a five point scale. The management want to see a report showing a table for each

shop comparing their percentage top box scores to the national average. The final table should only show the three best and three worst measures for each shop relative to the national average.

	National Avg.	ATP Cheese	Comparison to National Avg.
Highlights			
Friendly staff	*56%*	*61%*	***+5%***
Staff knowledgeable about cheese	*46%*	*50%*	***+4%***
Quality of produce	*42%*	*44%*	***+2%***
Lowlights			
Range of produce	*65%*	*59%*	***-6%***
Easy of parking	*58%*	*48%*	***-10%***
Appearance of shop	*46%*	*34%*	***-12%***

Figure 2 Example table based on manipulated, aggregated data

Logically, the process for producing this table is:

Aggregation

- Produce a table with the base and the distribution for all questions in the rows by all cheese shops in the columns.

Table manipulation

- Calculate the top box percentages by dividing the first element in the distribution by the base for each question.
- Calculate a derived column "National Avg." as the average for each question across all shops—the average needs to be shop based, not respondent based.
- Calculate a derived column "Comparison to National Avg." by subtracting the "National Avg." from the value for the current shop.
- Order the questions by the "Comparison to National Avg." column, descending.
- Select the top three and bottom three questions, the two derived columns and the current shop column.

Presentation

- Generate the table in word with the required labelling and cell formatting.

I have intentionally made this example a little awkward to do in a crosstab system but it is a fairly typical report and illustrates some of the most common table manipulations. Technically, there are many possible approaches to automating the process. Probably the most widely used would be to run initial aggregations in a crosstab system and then use Excel/VBA to perform the table manipulations.

In principle, there is nothing wrong with this technique, and for a one-off report, it is a practical solution. But requests for reports like this come up again and again with slight variations. Even the most skilled Excel programmer would have to admit that developing and maintaining the code needed to cope with the vast range of transformations needed for market research reporting is a non-trivial task.

One of the problems is that many of these transformations can only be generalised if the input table is thought of in more than two dimensions and many aspects of Excel are inherently two-dimensional. What we need is some flexible syntax for specifying transformations to multi-dimensional data tables in the way that SQL is used in the database field for transforming relational data tables. One possible candidate is MDX, a query language used across a range of OLAP database solutions to retrieve aggregated data from multi-dimensional "cubes".

4. OLAP and MDX

OLAP (OnLine Analytical Processing) systems are database systems designed specifically for holding aggregated data and have a whole data model and vocabulary of their own that needs to be mastered. The data is stored in structures known as “cubes” that can be thought of as multi-dimensional spreadsheets similar to an Excel pivot table. Each “dimension” of the cube is made up of a hierarchical collection of “members” which are similar to the elements within a categorical variable.

Going back to example 1, it is possible to build an OLAP cube to reproduce this table although you would never build a cube this small in practice. We’ll call the cube Ex1. It would have two dimensions—[Age] and [Gender]. Each dimension would have a default “all” member for the total and a member for each element of the variable. It is important to note that the dimension structure is hierarchical, the “all” member [All Gender] is a parent of [Gender].[Male].

Every cell in a cube can be uniquely identified by specifying one member from each dimension - this combination of members is called a “tuple”—for example the tuple ([Gender].[Male], [Age].[11-16]) identifies a single cell (value of 23).

If we want to get data out of the cube we need to run an MDX query, just as you would run an SQL query against a relational database. An MDX query has the same basic structure as an SQL query with a select clause to specify what you want, a from clause to say where you want to get it from. So, the most basic query would be:

```
select [Gender].members on rows, [Age].members on columns from ex1
```

The .members syntax returns all the members of a dimension. This query would return all the data in the cube. However, the cube only holds actuals. If we want to see column percentages as well we need to create a “calculated member” with the correct formula. The query for returning the full table with column percentages would be:

```
with
member [measures].ColPer as
 ‘ ( [measures].[actuals] / ( [measures].[actuals], [Gender].CurrentMember.Parent )) * 100 ‘
select
 [Gender].members on rows,
 [Age].members on columns,
 { [measures].[actuals] , [measures].[ColPer] } on pages
from ex1
```

This query illustrates a couple of important MDX concepts. Firstly, the query refers to a dimension called [measures]. This is an intrinsic dimension contained in all cubes. In research terms, if the other dimensions are like categorical variables, the members in the measures dimension are the numeric variables available as increments for the table. In this case the cube has been set up with one measure [actuals] which is just a straight count of the records. This was picked up as the default measure for the first query.

The next concept is the “with” clause which is used to set up any temporary definitions of calculated members. Here it contains the definition for a new calculated member to be added to the measures dimension called [measures].ColPer. The key part of this definition is the expression [Gender].CurrentMember.Parent.

When the query is run the query engine evaluates the whole expression in the context of every tuple it needs to display. The CurrentMember syntax can be used to refer to members within these iterations—it serves the same purpose as relative cell references in Excel, although in a very different way. The example is slightly confusing because CurrentMember references are implicit for dimensions not included in any tuples. The expression could be more explicitly written as:

```
(([measures].[actuals], [gender].CurrentMember, [age].CurrentMember ) /
([measures].[actuals], [gender].CurrentMember.Parent, [age].CurrentMember)) *100
```

This may seem like a lot of effort to produce just to produce simple column percentage but the advantage is that the syntax is general and has a rich set of functions that can be used to solve a very wide range of problems. To prove the point, the table manipulations required to produce the output table in figure 2 can all be expressed as a single, slightly daunting, MDX query, as follows:

```
With
/*create a member for the percentages*/
member [measures].ColPer as
 ‘ ( [measures].[actuals] / ( [measures].[actuals], [ratings].CurrentMember.Parent )) * 100 ‘,
solveorder = 1

/*create a member for the average percentage at a shop level*/
member [shop].[NationalAvg] as ‘ avg ( [Shop].[All Shop].children, [measures].ColPer ) ‘,
solveorder = 2

/*create a member comparing the current shop to the national average*/
member [shop].[Comparison] as ‘ [Shop].[ATP Cheese] - [Shop].[NationalAvg] ‘ , solveorder = 3

/*create a “top box” set by taking the first child of each question*/
set temp1 as ‘ [ratings].question.members ‘
set TopBoxes as ‘ generate (temp1, topcount ( [ratings].currentmember.children, 1) ) ‘

/*order by the Top Box set by the comparison member, descending*/
set temp2 as ‘ order (TopBoxes, ([measures].ColPer, [Shop].[Comparison]), bdesc) ‘

/*Create sets for the top 3 and bottom 3 questions
set Highlights as ‘ head (temp2, 3) ‘
set Lowlights as ‘ tail (temp2, 3) ‘

/*select the data
select
  { Highlights, LowLights }  on rows,
  { [Shop].[NationalAvg], [Shop].[ATP Cheese], [Shop].[Comparison] } on columns
from ex1
where ( [measures].[ColPer] )
```

This is not an MDX tutorial so I won’t go through this query or the cube it runs against in detail. I’ve included comments to show how closely the structure follows the logical description of the table manipulations required. This is one of the strengths of MDX. Once you are familiar with the syntax and functions available, it is usually possible to map a description of the problem directly to a series of MDX functions.

MDX tends to be easy to change. At a trivial level, I only need to substitute references to [ATP Cheese] with [NOP World of Cheese] to produce the next report, but, because queries are just a chain of functions acting on the basic data structures, dimensions, members, sets, and tuples, it is usually

possible to radically alter a section of a query without affecting the rest. This modularity also means you can almost always re-use parts of queries in different situations.

It may not be obvious from these examples but MDX is essentially a transformation language that translates an input cube into an output cube in the same way that SQL translates an input relational table into an output table. In a market research context this means that the output of each table transformation is just another table and queries can be linked together with the output from one forming the input to the next. Again, this can be very helpful in achieving modularity and code re-use.

One more advantage, and probably the most important from an application development perspective, MDX, like SQL, is a text based query language that is embedded into other applications. Going back to our - aggregation, table manipulation, presentation - model of table processing, it allows a complete separation of the table manipulation and presentation components by giving them a way of communicating. If you've ever tried to port any significant amount of Excel code to a web page (or any other application), you'll understand how important this can be.

If you want to know about the detailed mechanics of MDX I would recommend *MDX Solutions* by Thomsen and Spofford (2001). There is a wealth of information and detailed examples freely available on MDX, which is another strong argument for using it.

5. How can MDX be used on market research data?

I hope I've demonstrated how potentially useful a language like MDX could be in automating market research reports. The problem is that MDX is the language component of OLAP database systems. If the data hasn't been aggregated into an OLAP database you can't use MDX to manipulate it.

I should come clean at this stage and point out that I believe that full OLAP solutions are only useful in a very limited number of market research production scenarios. A full discussion of why I have come to this conclusion is outside the scope of this paper but my personal view could be summarised as follows: although OLAP can in theory meet most reporting requirements, the overhead involved in setting-up and maintaining the complex OLAP databases needed for market research data is far too high for a typical agency production environment.

The solution to this problem is to bypass the aggregation to an OLAP database by developing a table manipulation engine that understands individual market research tables as very small cubes and can therefore apply MDX transformations to them. This would then be available to any application as a service on a server just like any other database system.

6. Crosstabs as Cubes

It is easy to think of all tables as small cubes with three dimensions, [Rows], [Columns], and [Measures], but to be able to run effective MDX against the tables it is important that they can be loaded with as much structure and meta-data as possible.

For example, even a simple table such as the one shown in figure 3 presents a few problems if we are just loading from a text file:

Q1—Whether ever buy cheese

Base: All respondents

		Sex		Age					Blue Cheese Consumer	
	Total	Male	Female	11-20	21-34	35-54	55+	Mean	Yes	No
Unweighted Base	**908**	439	469	19	96	159	175	37.3	562	346
Weighted Base	**1000**	483	517	21	106	175	193	38.5	619	381
Yes	**813**	393	420	19	84	144	159	39.2	508	305
No	**187**	90	97	2	22	31	34	38.2	111	76

Figure 3 Example table with a hierarchical structure

The biggest problem is that it is not clear what hierarchical structure of the row and column dimensions should be. Some of the most useful MDX functions are designed to navigate hierarchical structures. If we want to take advantage of them we need to understand the axes as hierarchies. At the very least we need to load elements as children of their bases so we can use the ".parent" function to calculate percentages (see the examples above).

There are other MDX functions that are clearly hierarchical like Cousin, FirstChild, and FirstSibling, but there are also more subtle reasons to understand the structure as a hierarchy. For example, the order function has an option that allows you to preserve or break the hierarchy. By default it will preserve the hierarchy and only order elements within their parent, which is what is usually required.

For this table I would load three levels for the columns: total, variables and elements. So the full position of the 55+ element in the hierarchy would be [columns].[total].[age].[55+]. To make it useful to use the .parent function on this element, I would also replicate the data for [columns].[total] in [columns].[total].[sex], [columns].[total].[age], [columns].[total].[blue cheese consumer].

Another problem is that you could consider this table, and practically all market research tables, as an untidy, partial projection of a multi-dimensional table. In an ideal world, this table would probably have four members in the [Measures] dimension: [Measures].[Unweighted]—a full table of unweighted data, [Measures].[Weighted]—most of the table shown, [Measures].[UnwtMean] - a table of unweighted means, and [Measures].[WtdMean]—a table of weighted means.You could load the data in this structure and just leave most of the dimensions 2 to 4 empty.

Another more pragmatic approach is to load all the data into one [measure] and make sure the members are tagged with enough meta-data to make sensible selections when running queries. For example, it is easy to construct MDX queries that exclude the statistical elements from percentages or sorting as long as the members have been loaded with a member property to identify them. It would only make sense to load data into an additional member in the measures dimension when there is a more or less complete set of clearly identifiable data, such as row and column percentages.

To sum up, yes, it's possible to understand crosstabs as small cubes and load them into an MDX based manipulation engine but how well this works depends on the structure of the source and how much meta-data it provides to the engine. The crucial requirements are that the engine can understand the hierarchical structure of the table and has enough information to construct a meaningful measures dimension.

7. Table Sources

Text based table files can used as a source for this process but they will need to be go through a complex parsing process to identify the different elements of the table. An application like ATP's WinYaps which can parse a table set and then export a more structured representation of the file, makes it possible to bind onto traditional table set as a source.

Any moves towards common standards for tables will greatly simplify the process provided the standard supports the structures required. The XML table structure proposed by the Open Survey group provides a perfectly adequate table source.

It is also likely that any future desktop crosstab applications will expose representations of tables either as published object models or XML file structure that will make this binding process very easy and much more reliable. For example, I anticipate that it will be very easy to use tables created in SPSS's MRTables as a source.

Another possible source is XMLA, "an open industry-standard web service interface designed specifically for online analytical processing." The XMLA specification, among other things, defines an XML structure for tables that is based on an OLAP/MDX concept of data and is therefore ideally suited to the problem. The web-service component of XMLA also opens up some exciting possibilities. The vision is that Business Intelligence applications will be split into front-end XMLA consumers and back end databases that act as XMLA provides. In theory, this means that if you developed a table manipulation engine as an XMLA provider, any XMLA consumer (Excel, Cognos, Business Objects) would be able to access the table data though XMLA using MDX as the language component of the interface.

It also means that it would be possible to incorporate references to external data held in other XMLA providers inside table based queries. One example would be incrementing a market research table by actual sales data from a companies internal BI system—or the other way round.

References and Further Reading

Thomsen E & Spofford G (2001) MDX Solutions: with Microsoft SQL Server Analysis Services, John Wiley & Sons

XMLA web site See www.xmla.org.

About the Author

John Lyon has more than 17 years experience of developing information delivery systems while at NOP. He has developed PC systems, client-server systems, OLAP systems and web based systems, and has seen the evolution of electronic information delivery within the market research industry. John worked at NOP until 2001 where he was Director of IT Development. He is currently a Senior Consultant within the ATP IT Solutions Division and is spearheading the development of information delivery systems, including Vector.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

triple-s. Managing Success

Geoff Wright

Abstract

As developers of software and applications, we all want the results of our work to be widely adopted and used. But in the ever-changing environment of computing this type of success can lead to problems. The triple-s standard for survey data interchange has steadily become more successful since it was first published 10 years ago. This paper attempts to elucidate these problems and describes some steps that the triple-s Group are working on to ensure the continued success of triple-s.

Keywords

Standards; Survey Interchange; Survey Metadata; triple-s; XML

1. Background

The idea of a simple mechanism for overcoming the proliferation of bespoke imports and exports between different survey packages came from Peter Wills of Mercator in a paper given at the First ASC International Conference in Bristol (Wills, 1992). The triple-s survey interchange standard was developed in response to this need and has steadily become more successful since the first version was published nearly 10 years ago.

The standard defines two text files that describe the survey data. The *Definition File* contains general information about the survey together with descriptions of the survey variables (i.e. the metadata). The *Data File* contains the actual case data for the survey. Although many consider the definition file to be the triple-s standard, it is in fact the formal definition of both files that constitutes the standard (i.e. triple-s is not a means for describing an arbitrary data file).

2. Success has its problems

Following the development of the first XML version of the standard (Hughes et al, 1999) and its publication in 2000, there are now more than 25 software suppliers who support the standard and have provided facilities that allow the user to import or export triple-s files. But this success has produced problems that we did not fully anticipate when we started:

- Increased expectation of what survey information (i.e. the metadata) should be made available.
- Increased complexity of survey structures (e.g. hierarchical).
- Use of survey data outside the traditional data analysis packages (e.g. bespoke VB programs).
- General developments in computing (e.g. XML, Unicode).

Addressing these issues has lead to steady development of the standard, producing a number of different versions (1994: original version 1.0, 1996: version 1.1, 2000: XML version 1.1, and 2002: XML version 1.2). With the exception of the move to XML for the definition file, none of these enhancements has been radical. But even so, it has been difficult to expect all software suppliers to continually keep up with these changes. This has produced the situation where it is no longer necessary to just ask if the package supports triple-s, but now the user needs to know which versions. We are in danger of asking suppliers to replace their proliferation of bespoke data transfer modules with an increasing number of triple-s import/export facilities!

Existing implementers

The existing implementers of the earlier triple-s standards are the most affected by the problem of multiple versions. They will have already committed resources to initially supporting triple-s, and it could be difficult to justify more work for what may only be a small increase in functionality. Because some members of the triple-s Group are involved in the development of survey analysis software, we understand fully this dilemma.

With the XML version 1.2 standard we have tried to make the transition to a new version easier by ensuring that the new standard accept the previous standard as a valid subset. Also an advantage of XML is that it should be more straightforward for an importer to ignore some elements and attributes. So, for example, an importer for XML version 1.2 should also be able to process an XML version 1.1 definition, and an existing importer of XML version 1.1 should only have to make minor changes to process an XML version 1.2 definition ignoring the new features.

As another partial solution to the problem and also to encourage users to move forward, the triple-s Group have provided a simple converter that takes a ‘classic’ triple-s definition and produces the XML format. This at least helps users who are using export software that has not moved on the XML versions of the standard. Providing such a converter that goes ‘forward’ is not trivial, but is at least straightforward in concept. However a converter that goes ‘backward’ would be a more substantial undertaking because each new version introduced new features that may be difficult or impossible to reproduce in an earlier version of the standard.

New implementations

A second consequence of the increased use of triple-s as the format of choice when moving survey data around is the pressure of users on software suppliers. For those deciding to support triple-s now we feel the need to provide extra assistance beyond the published standard.

For these suppliers we can provide help and advice in selecting a suitable XML parser. We are also willing to use our experience to assist in interpreting the triple-s standard. In addition the triple-s Group have been working on providing a suite of test definition and data files for checking triple-s import procedures. Some work has also been done on producing a validation utility for checking exported triple-s files.

Non implementers

The third group is those suppliers who have chosen not to implement the standard. As the standard becomes more widespread these become more of a problem. The pressure of users can be a significant influence in getting suppliers to provide what is in reality a relatively straightforward piece of

development. But in some cases the package is no longer being developed, or the supplier provides their own 'open' access to the survey data.

In a number of these cases we have welcomed the involvement of third-party companies and the support of the ASC in providing triple-s import/export capabilities without requiring any development from the original developers. This approach is obviously less satisfactory than a native import/export provided and supported by the supplier, but the modules that have been provided so far on our website (http://www.triple-s.org) have seen a steady number of downloads.

3. Moving the standard on

But we need to move the triple-s standard on again. In particular the description of hierarchical survey data needs to be addressed, and extending the data file format to provide more than just fixed format fields. This will inevitably produce yet another version of the standard, and will raise the barrier for both existing implementers and those who have not yet implemented triple-s support.

As a consequence we feel that the triple-s Group must now do more than just produce the triple-s XML definition of the standard. At the minimum we should enhance our conversion utility, provide assistance for those upgrading to the new version of the standard, and try to produce test definitions and validation programs to assist developers. However we see the most important way in which we can reduce the 'versions' problem and make it easier for developers to implement support for triple-s, is to work on producing an API (Application Program Interface) for triple-s. The objective would be to allow developers to remove themselves from the details of the triple-s XML definition and try to isolate them from future changes in the standard.

Of course not everyone will want nor need to adopt a triple-s API. We expect that existing implementers of triple-s XML would be likely to postpone any change until they wanted to support a future version of triple-s. But new implementers would be encouraged to use the API from the start.

Outline of API requirements

Producing an API for triple-s is a very different task to producing the triple-s standard itself. The aim is that the definitive standard will remain the description of the files—the API would only be a 'recommended' way to access the information in those files. But there are a number of factors that determine what this 'access' should be.

- Languages and platforms to support
- Overall scope of an API
- Design of the API
- Limitations of an API

Languages and platforms

Survey software today is written in a variety of languages and runs on a number of operating systems. It would be unrealistic to expect that we could support all combinations. Also we need to interface to one or more existing XML parsers which may limit our options. At this stage we anticipate that C++ is probably the most universal implementation language, and that COM will be the most useful implementation for Windows users. Possibly the most acceptable approach would be to define the API in terms of the Interface Definition Language defined by the Object Management Group (ISO standard 14750), which has defined mappings to many major programming languages.

Overall scope of an API

Different types of user will require different functionality from an API. In particular, the API should cover both the import and export of triple-s surveys. Survey packages that already process fixed format character data may only need information from the Definition File because they can use their existing mechanisms to handle access to the Data File. However, other users may prefer to access the case data via the API, hence devolving responsibility for locating and interpreting the actual data values.

Design of the API

We are already aware of existing APIs in this general area, but the objective of simplicity that underlies triple-s may mean that we have to risk the accusation of "reinvention". Our general view is that the API will be most universal if it only returns 'atomic' values (i.e. int, float, character, string). This does not preclude a higher level or more object oriented model, which could be built on top of this basic API.

Limitations of an API

In general the triple-s standard does not impose limits (e.g. number of characters in a text, digits in a number). But access through an API will of necessity introduce some limits. For example, if the lower and upper values of an integer range are returned as 'int' then there is an implicit limit on these values. However, we anticipate that such implicit limits will be less than those within the survey software itself.

A second problem lies in the effect of character 'encoding'. XML itself is Unicode, but many languages do not have 'atomic' values for returning items like Unicode strings.

4. Conclusion

We see the provision of an API as an important step in ensuring the continued success of triple-s. With the possibility of a new standard that supports hierarchical data, one aim will be to make it easier for products to support different versions of the standard. A second objective is to allow developers to remove themselves from the details of the triple-s XML definition. In order to make an API useful and useable, we intend to draw heavily on the experiences of the existing implementers of triple-s.

References

Hughes K, Jenkins S and Wright G (1999) triple-s XML: A Standard Within a Standard, Proceedings of the 3rd ASC International Conference. Edinburgh UK.

Wills P (1992) Data Use and Reuse: The movement and use of survey data across different hardware and software platforms, Proceedings of the 1st ASC International Conference. Bristol UK.

About the Author

Geoff Wright is a director of Computable Functions Ltd., an independent survey and media research IT consultancy. He has been using, designing, and advising on survey and statistical software for more than 30 years, and has been a member of the triple-s Group since its inception. He can be contacted at Computable Functions Ltd., Enigma House, 30 Alan Turing Road, Guildford GU2 7AA, England; e-mail: geoff@computable-functions.com.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Automatic Coding of Open-Ended Surveys Using Text Categorization Techniques

Daniela Giorgetti, Irina Prodanof, Fabrizio Sebastiani

Abstract

Open-ended questions do not limit respondents' answers in terms of linguistic form and semantic content, but bring about severe problems in terms of cost and speed, since their coding requires trained professionals to manually identify and tag meaningful text segments. To overcome these problems, a few automatic approaches have been proposed in the past, some based on matching the answer with textual descriptions of the codes, others based on manually building rules that check the answer for the presence or absence of code-revealing words. While the former approach is rarely effective, the major drawback of the latter approach is that the rules need to be developed manually, and before the actual observation of text data. We propose a new approach, inspired by work in information retrieval (IR) that overcomes these drawbacks. In this approach survey coding is viewed as a task of *multiclass text categorization* (MTC), and is tackled through techniques originally developed in the field of *supervised machine learning*. In MTC each text belonging to a given corpus has to be classified into exactly one from a set of predefined categories. In the supervised machine learning approach to MTC, a set of categorization rules is built *automatically* by learning the characteristics that a text should have in order to be classified under a given category. Such characteristics are automatically learnt from a set of training examples, i.e. a set of texts whose category is known.

For survey coding, we equate the set of codes with categories, and all the collected answers to a given question with texts. Two of the paper's authors have carried out automatic coding experiments with two different supervised learning techniques, one based on a naïve Bayesian method and the other based on multiclass support vector machines. Experiments have been run on a corpus of social surveys carried out by the National Opinion Research Center, University of Chicago (NORC). These experiments show that our methods outperform, in terms of accuracy, previous automated methods tested on the same corpus.

Keywords

Open-ended survey coding, multiclass text categorization, machine learning, information retrieval.

1. Introduction

The survey process typically consists of several steps which have to be performed sequentially, and include the design, construction, administration, analysis and evaluation of the survey itself. Survey questions may be open-ended, i.e. they allow respondents to answer in their own words, or close-ended, i.e. they consist of questions where the answers are limited by a multiple choice list or entail a

true/false, agree/disagree reply. In this paper we will only deal with coding, a core step in open-ended surveys, in which text data extracted from free responses to open-ended questions are mapped into codes according to a predefined code scheme. While the coding of a close-ended survey is straightforward, the coding of an open-ended survey is a subjective task traditionally performed manually by trained professionals, and thus it is costly, time-consuming and affected by the inter-coder agreement problem (which occurs when there is no agreement between two different coders on the code to be assigned to a specific text segment). These are the reasons why social scientists tend to avoid including too many open-ended questions in their surveys, preferring the less problematic multiple choice questions.

A few solutions have been proposed to automate the coding phase of open-ended surveys, but they still require some manual engineering effort. Our approach instead, relies on fully automatic text categorization techniques developed in the information retrieval and machine learning fields.

The structure of this paper is as follows. In Section 2 we focus on the task of automatic survey coding and we discuss how previous approaches have tried to automate (at least partially) the coding task. In order to make the paper self-contained, in Section 3 we give some background on information retrieval and text categorization. In Section 4 we describe how we map the task of coding open-ended surveys onto a text categorization task. In Section 5 experimental results obtained (Giorgetti and Sebastiani (2003)) on three corpora from NORC General Social Survey using two different text categorization classifiers are presented, and in Section 6 we summarize our approach and the results achieved and give some hints for possible future developments.

2. Automatic coding of open-ended surveys

Coding a survey may be viewed as the task of identifying and labelling common meaningful themes in the answers given by different respondents to a survey questionnaire. Sometimes novel, unforeseen themes are sought in the answers being analyzed, but most of the times surveys are run with a clear purpose, i.e. with a clear set of previously identified concepts whose presence in the text corpus must be assessed or measured in some way. We will concentrate on the most frequent case, in which themes sought are known a priori.

The issue of automating survey coding has been dealt with in software packages developed in the context of text analysis for the social sciences, which includes survey analysis. However, these software packages are not usually tailored for the specific task of survey analysis; in particular, the solutions they provide for the coding task are still unsatisfactory. Many of these applications mostly aim at facilitating the manual coding of data, and their display in several convenient ways. A few of these packages instead do perform automatic coding, mainly by relying on specialized hand-crafted *dictionaries* (or *rules*). This means that text fragments are automatically assigned to a specific category if and only if they contain words matching those in the dictionary relevant to the category.

The closest approach to ours is probably the dictionary-based approach as it is described in Viechnicki's work (1998). In this paper, responses to questions from the NORC[1] General Social Survey (Davis and Smith (1996)) are classified by means of a set of codes predefined by NORC social scientists. Viechnicki proposes two alternative approaches. In the first one, words that characterize a given category may be combined by means of boolean operators (AND, OR, NOT), and the answer is

[1] at http: //www.norc.uchicago.edu/

classified under the category whose boolean description it matches. The second method is instead based on computing the similarity between two weighted vectors of words (see subsection Document indexing of Section 3 for an explanation of how weighted vectors of words and their similarity are computed) extracted from the answer and from a textual explicatory caption of the code, and choosing the code with the highest similarity score. A variant of a dictionary-based automated approach is proposed by Macchia and Murgia (2002), in which the answer is assigned a unique code if there is an exact match with phrases belonging to a previously defined dictionary associated with the code, or it is assigned a “best code" if the match is partial.

These approaches have the typical drawback of dictionary-based methods, which need a dictionary to be manually developed before the actual coding step takes place, i.e. when data is still totally unknown. Besides, dictionaries need to be manually updated as a result of changes in the structure and/or semantics of the coding scheme. For example, if a new category is added to the scheme, a new dictionary has to be created for it, and the dictionaries for the old categories need to be updated in order to avoid “capturing” the answers that are instead to be filed under the newly introduced category.

Our learning-based approach to survey coding has several advantages with respect to the dictionary-based approach.

Firstly, the manual effort is directed towards the manual coding of a relatively small training set of answers, and not towards the creation of specialized dictionaries. This is advantageous, as it is easier to manually classify a set of documents than to build and tune a dictionary of words that trigger the attribution of the code, for the simple fact that it is easier to characterize a concept extensionally (i.e. to select instances of it) than intensionally (i.e. to describe the concept in words, or to describe a procedure for recognizing its instances).

Secondly, it is solidly grounded in machine learning theory, and it can leverage on a wealth of results and techniques developed within text categorization, a discipline which has been bursting with activity in the last ten years (see e.g. Sebastiani (2002)) and has produced systems whose accuracy rivals or exceeds that of a human (i.e. systems capable of generating codes that correlate with those attributed by a coder at least as well as the codes attributed by two human coders correlate with each other).

Of course, our approach is mostly useful for medium- to large-sized surveys, as in the learning phase we need a hand-coded set of answers to train the inductive learner. This means that if a survey is somewhat limited in the number of surveyed people, hand-coding the training set may coincide with hand-coding the entire set. NORC’s surveys are examples of relatively large-sized surveys, since 40,933 interviews have been completed in the years from 1972 to 2000.

3. Basics of Text Categorization

This section gives a brief introduction to text categorization and to some basic concepts adopted from the related field of information retrieval; these notions will provide us with the background for solving the survey coding task.

Text categorization (TC, also known as *text classification*) is the task of approximating the unknown *target function* $\Phi : D \times C \rightarrow \{T, F\}$ that describes how documents ought to be classified by means of a function $\hat{\Phi} : D \times C \rightarrow \{T, F\}$ called the *classifier*, where $C = \{c_1, ..., c_{|C|}\}$ is a predefined set of thematic categories, D is a domain of documents, and T, F represent the boolean values *true* and *false*. If $\Phi(d_j, c_i) = T$, then d_j is called a *positive example* (or a *member*) of c_i, while if

$\Phi(d_j, c_i) = F$ it is called a *negative example* of c_i. The categories are just symbolic labels, and no additional knowledge of their meaning is usually available. Classification has to be accomplished only on the basis of knowledge extracted from the document contents, as no metadata (such as e.g. publication date, document type, publication source) are usually available.

Text categorization is a *subjective* task: when two experts (human or artificial) decide whether to classify document d_j under category c_i, they may disagree, and this in fact happens quite frequently. A news article on Tony Blair visiting Tuscany could be filed under Politics, or under Gossip, or under both, or even under neither, depending on the subjective judgment of the expert.

In some applications each $d_j \in D$ must be assigned a unique label $c_i \in C$, while in others each $d_j \in D$ may be assigned any number $0 \leq n_j \leq |C|$ of categories. The former case is usually dubbed the *binary* case or the *multiclass* case, depending on whether $|C| = 2$ or $|C| > 2$, respectively. Since the $|C| > 2$ case will be the object of interest in this paper, from here on when speaking of TC we will actually mean *multiclass* TC.

We can roughly distinguish three different phases in the life cycle of a TC system: document indexing, classifier learning, and classifier evaluation. The three following paragraphs are devoted to these three phases, respectively; for a more detailed treatment see Sections 5, 6 and 7, respectively, of Sebastiani (2002).

Document indexing

Document indexing denotes the activity of mapping a document d_j into a compact representation of its content that can be directly interpreted (i) by a classifier-building algorithm and (ii) by a classifier, once it has been built. The document indexing methods usually employed in TC are borrowed from IR, where a text d_j is typically represented as a vector of term *weights* $\vec{d}_j = \langle w_{1j}, \ldots, w_{|T|j} \rangle$. Here, T is the *dictionary*, i.e. the set of *terms* (also known as *features*) that occur at least once in at least α documents, and $0 \leq w_{kj} \leq 1$ quantifies the importance of t_k in characterizing the semantics of d_j. Typical values of α are between 1 and 5.

An indexing method is characterized by (i) a definition of what a term is, and (ii) a method to compute term weights. Concerning (i), the most frequent choice is to identify terms either with the *words* occurring in the documents (with the exception of *stop words*, i.e. topic-neutral words such as articles and prepositions, which are eliminated in a pre-processing phase), or with their *stems* (i.e. their morphological roots, obtained by applying a stemming algorithm). When terms coincide with words (either stemmed or not) the approach to document representation is called *bag of words*. Experimental results (Fuhr et al. (1991); Schütze, Hull and Pedersen (1995); Tzeras and Hartmann (1993)) on the use of *phrases* as additional indexing terms have not been uniformly encouraging up to now. Lewis (1992) argues that the likely reason for the disappointing results is that, although indexing languages based on phrases have superior semantic qualities, they have inferior statistical qualities with respect to word-only indexing languages.

Concerning (ii), either statistical or probabilistic techniques are used to compute terms weights, the former being the most common option. Binary weights are a special case where 1 denotes the presence and 0 the absence of the term in the document. The choice of binary or non-binary weights depends on the kind of input that the classifier learning algorithm adopted requires. A popular class of statistical term weighting functions yielding non binary weights is *tfidf* (see e.g. Salton and Buckley (1988)),

where two intuitions are at play: (a) the more frequently t_k occurs in d_j, the more important for d_j it is (the *term frequency* intuition); (b) the more documents t_k occurs in, the less discriminating it is, i.e. the smaller its contribution is in characterizing the semantics of a document in which it occurs (the *inverse document frequency* intuition). Weights computed by $tfidf$ are often normalized so as to contrast the tendency of $tfidf$ to emphasize long documents. Note that formulae such as $tfidf$ weigh the importance of a term to a document in terms of occurrence considerations only, disregarding for example the order in which terms appear in the document and their syntactic role.

In TC, unlike in IR, a *dimensionality reduction* phase is often applied so as to reduce the size of the document representations from T to a much smaller, predefined number. This has both the effect of reducing *overfitting* (i.e. the tendency of the classifier to better classify the data it has been trained on with respect to new unseen data), and to make the problem more manageable for the learning method, since many such methods are known not to scale well to high problem sizes. Dimensionality reduction often takes the form of *feature selection*: each term is scored by means of a scoring function that captures its degree of (positive, and sometimes also negative) correlation with c_i, and only the highest scoring terms are used for document representation.

Classifier learning

A text classifier for categories $C = \{c_1,...,c_{|C|}\}$ is automatically generated by a general inductive process (the *learner*) which, by observing the characteristics of a set of documents pre-classified under C gleans the characteristics that a new unseen document should have in order to belong to a generic category $c_i \in C$. In order to build classifiers for C one thus needs a labelled corpus Ω of documents such that the value of $\Phi(d_j, c_i)$ is known for every $\langle d_j, c_i \rangle \in \Omega \times C$. In experimental TC it is customary to partition Ω into three disjoint sets: Tr (the *training set*), Va (the *validation set*), and Te (the *test set*). The training set consists of documents the learner "observes" to build up the classifier. This is called a *supervised* learning activity, since learning is "supervised" by the information on the membership of training documents in categories. The validation set is the set of documents the engineer uses to fine-tune the classifier, e.g. choosing for a parameter p on which the classifier depends, the value that has yielded the best effectiveness when evaluated on Va. The test set is used for the final evaluation of classifier effectiveness. In both the validation and test phase, "evaluating the effectiveness" means running the classifier on a set of pre-classified documents (Va or Te) and checking the degree of correspondence between the output of the classifier and the pre-assigned labels (which are assumed to be correct).

Several methods have been proposed in the text categorization literature for learning a text classifier from training data (see Sebastiani (2002) for a review), including probabilistic methods, regression methods, decision tree and decision rule learners, neural networks, batch and incremental learners of linear classifiers, example-based methods, support vector machines, genetic algorithms, hidden Markov models, and classifier committees.

Classifier evaluation

Training efficiency (i.e. average time required to build a classifier $\hat{\Phi}$ from a given corpus Ω), as well as *classification efficiency* (i.e. average time required to classify a new document by means of $\hat{\Phi}$), and *effectiveness* (i.e. average correctness of $\hat{\Phi}$'s classification behaviour) are different measures of success for a learner. However, effectiveness is usually considered the most important criterion, since

in most applications one is willing to trade training time and classification time for correct decisions. Also, it is the most reliable one when it comes to comparing different learners, since efficiency depends on too volatile parameters (e.g. different software/hardware platforms). As a result, we will only measure the success of our approach in terms of effectiveness.

In multiclass TC, effectiveness is usually equated to *accuracy*, which is defined as the ratio between the number of correct decisions and the total number of decisions.

4. Automatic coding of open-ended surveys by a Text Categorization approach

In this section we describe how survey coding may be handled as a TC task, i.e., as the task of automatically generating a classifier that automatically selects, from a set of predefined codes, the correct code to attach to a given answer. All the answers to a given question *q* play the role of domain D, and all the possible codes that may be attributed to an answer to question *q* play the role of the set of categories C (coding answers to different questions thus corresponds to different TC tasks).

The input to the learners (and to the classifiers, once they have been built), consists of a set of answers d_j represented as vectors of term weights $\vec{d}_j = \langle w_{1j},...,w_{|T|j} \rangle$. Note that the fact that many of the answers in the corpus are ill-formed (like in the sentence "my boyfriend went to court yestreday (sic) and if he doesn't have a drivers liceces (sic) insurance by the 28th of this month he goes to jail for 6 months", extracted from the NORC corpus *angry_at* described in Section 5) makes the bag of words approach to representation even more appropriate: given that current linguistic analysis techniques have not proven worthy (i.e. well-performing and robust at the same time) in standard TC, where we usually deal with syntactically well-formed text, it is easy to conjecture that they could hardly prove worthy here.

Giorgetti and Sebastiani (2003) have run a series of experiments with two different classifier-learning methods. The first learner used is a probabilistic naïve Bayesian learner (McCallum and Nigam (1998)), as implemented in the RAINBOW package[1] . Probabilistic text classification methods assume that the data was generated by a parametric model, and use the training set to estimate the parameters of this model. Bayes' theorem then allows to estimate from this model the probability that a given category has generated the document to be classified; classification thus consists in selecting the category with the highest probability. The second learning method used is a multiclass *support vector machine* (SVM) learner as embodied in the MCSVM software[2] . SVMs attempt to learn a hyperplane in $|T|$-dimensional space that separates the positive training examples of category c_i from the negative ones with the maximum possible margin, i.e. such that the minimal distance between the hyperplane and a training example is maximum; results in computational learning theory indicate that this tends to minimize the generalization error, i.e. the error of the resulting classifier on yet unseen examples. SVMs were initially conceived for solving binary classification problems, and only recently they have been adapted to multiclass classification.

[1] RAINBOW was implemented by Andrew McCallum and can be downloaded from http://www.cs.cmu.edu/~mccallum/rainbow.

[2] MCSVM was implemented by Koby Crammer and Yoram Singer, and we were kindly provided with a pre-release version of it.

Crammer and Singer (2001) describe an algorithmic implementation of multiclass SVMs based on a notion of margin generalized to multiclass problems, which allows to train directly a multiclass classifier (while in most of previous work the multiclass problem is decomposed into multiple independent binary classification tasks as in Hsu and Lin (2002)).

Regarding effectiveness, the text categorization literature has shown that naïve Bayesian approaches are, with respect to other learning methods, no more than average performers (see e.g. Dumais et al. (1998); Joachims (1998); Li and Yamanishi (2002); Yang and Liu (1999)). On the contrary, support vector machines are currently (together with "boosting"-based classifier committees) the unsurpassed top performers in the TC field (Dumais et al. (1998); Joachims (1998)). The reason why we experimented with RAINBOW is that we wanted to show that a text categorization approach to survey coding is much more effective than the dictionary-based approach *regardless of the specific learning method adopted*, i.e. that even with an average-performing learning method our text categorization approach to survey coding can outperform the dictionary-based method. Instead, the reason why we experiment with MCSVM is that we want to show what level of effectiveness this approach can achieve, once instantiated with a top-performing learning algorithm[1] .

A binary representation as input to RAINBOW, and a non-binary one as input to MCSVM have been adopted. This is due to the fact that the probabilistic models upon which RAINBOW is based require binary inputs, while this is not the case for SVMs. In the binary representation, w_{kj} represents just presence or absence of term t_k in answer d_j. Our non-binary representation is instead the *tfidf* function in its standard "ltc" variant (Salton and Buckley (1988)), i.e.

$$tfidf(t_k, d_j) = tf(t_k, d_j) \cdot \log \frac{|Tr|}{\#Tr(t_k)} \qquad (1)$$

where $\#Tr(t_k)$ denotes the number of answers in the training set Tr in which t_k occurs at least α times and

$$tf(t_k, d_j) = \begin{cases} 1 + \log \#(t_k, d_j) & if\ \#(t_k, d_j) > 0 \\ 0 & otherwise \end{cases}$$

where $\#(t_k, d_j)$ denotes the number of times t_k occurs in answer d_j. Weights obtained by Equation Table 2 are normalized by cosine normalization, yielding

$$w_{kj} = \frac{tfidf(t_k, d_j)}{\sqrt{\sum_{s=1}^{T} tfidf(t_s, d_j)^2}} \qquad (2)$$

In all the experiments, stop words, punctuation, and numbers, have been removed, and all letters have been converted to lowercase. No feature selection has been performed. The reason is that, as shown in extensive experiments by Brank et al. (2002), the effectiveness of SVMs is usually worsened by feature selection, irrespectively of the feature selection algorithm used and of the chosen reduction

[1] Note that there are no published results yet concerning the application to TC of *multiclass* SVMs, because multiclass SVMs are a recent development, and because most TC applications are binary. The assumption that multiclass SVMs would be a top performer once used in a multiclass TC context is based on the top performance that multiclass SVMs have delivered in multiclass application contexts other than TC (Crammer and Singer, 2001).

factor (this is also independently confirmed by the results of Taira and Haruno, (1999)), and the effectiveness of naïve Bayesian methods does not show systematic patterns of improvement either.

5. Experimental settings and results

As already pointed out, the experiments have been carried out on data from NORC's General Social Survey. This survey, which is ongoing since 1972, aims at investigating how people assess their physical and mental health, the balancing of security and civil liberties, external and internal security threats, inter-group relations and cultural pluralism, religious congregations, etc. We deal with three datasets (see Table 1) from the NORC General Social Survey administered in 1996. Each of these datasets (here nicknamed *angry_at*, *angry_why*, and *brkdhlp*) consists of a set of answers to a given question, plus their associated category codes manually chosen by NORC's professional coders from a predefined set of category codes[1]. The task consists in choosing exactly one code for each answer.

Table 1: Distribution of the categories in the three datasets used in the experiments.

Dataset	Category	No. of instances	Dataset	Category	No. of instances	Dataset	Category	No. of instances
angry_at	ANGRYFAM	275	*angry_why*	SELF	29	*brkdhlp*	FAMILY	57
	ANGRYWRK	345		PREVENTED	36		FRIEND	33
	ANGRYGVT	74		CRITICAL	88		GROUP	2
	WRK&GVT	8		DEMANDING	60		CLERGY	55
	WRK&FAM	27		EXPECT	196		PSYCHYAT.	56
	FAM&GVT	16		OTHER	51		AGENCY	16
	OTHER	625					OTHER	148
	total	1370		*Total*	460		*total*	367

We have chosen these three datasets because they are the same datasets used in Viechnicki (1998), which means that we will be able to obtain a direct comparison between the effectiveness of a method representative of the dictionary-based approach to survey coding and the effectiveness of our supervised learning approach.

For each dataset, the experiments were carried out according to the following main steps:

1. pre-process the data in order to obtain an input data format compatible with the learners (this had to be repeated once for RAINBOW and once for MCSVM, since the two systems require different data representations);
2. partition the set of answers in each dataset in four random disjoint subsets of equal size;
3. run the learner to generate a classifier, using three of the four subsets as the training set (75% of data) and the fourth as the test set (25% of data);
4. run the classifier to categorize the data in each test set of each dataset and evaluate the results in terms of accuracy.

In order to achieve better statistical significance, in all experiments steps 3 and 4 were repeated four times, for all four possible choices of the test set. Each of the results we report is thus the result of averaging across four different experiments. Accuracy has been computed on the three datasets both

[1] The *angry_at* and *angry_why* datasets actually involve the same question, which deals with the description of a situation that caused anger to the respondent; each answer was classified according to *two* different sets of codes, one concerning the object of anger, the other concerning the cause of anger. Actually, *angry_why* contains only a subset of the answers contained in *angry_at*. The *brkdhlp* dataset consists of answers to the question as to what source of help was used to deal with a nervous breakdown.

with RAINBOW and with MCSVM; the results are reported in Table 2, where they are compared with the accuracy obtained in Viechnicki (1998)on the same datasets.

Table 2: Comparative accuracy

	Dictionary-Based		Supervised Learning	
	Vector	Boolean	RAINBOW	MCSVM
angry_at	0.451	0.465	0.714 (+54%)	**0.756 (+63%)**
angry_why	0.211	0.272	**0.389 (+43%)**	0.376 (+38%)
brkdhlp	0.646	**0.747**	0.653 (-13%)	0.746 (-0.13%)
Average	0.436	0.495	0.585 (+18%)	**0.626 (+26%)**
Std. Dev.	0.218	0.239	**0.173 (-28%)**	0.216 (-10%)

Comparative accuracy (defined as the percentage of classification decisions that are actually correct, see also end of Section 3) results obtained on the *angry_at*, *angry_why* and *brkdhlp* datasets using a boolean and a vector-based method and using a naïve Bayesian and a multiclass SVM TC method. The percentile improvements in accuracy and average accuracy, and the percentile reductions in standard deviation, are reported with respect to the boolean method, the best dictionary-based method in Viechnicki (1998). **Boldface** indicates the best performance on the dataset.

The first observation we can make is that the supervised learning approach to survey coding significantly outperforms the dictionary-based approach: the improvements with respect to the best-performing method reported in Viechnicki (1998) are significant, a +18% on average for RAINBOW and a +26 % for MCSVM. The improvement is especially noteworthy on the "non-obvious" datasets: for instance, *angry_why* appears to be a hard to characterize dataset, as shown by the poor performance of the two dictionary-based methods, and on this dataset the supervised learning methods improve up to +43% with respect to them. The *angry_at* dataset looks somehow "easier" than *angry_why*, as witnessed by the fact that all four methods listed in Table 2 perform better on *angry_at* than on *angry_why*. This might also be explained by the fact that, as it can be seen from Table 1, it contains more data, since each category in *angry_at* has 195 positive examples on average, while this goes down to 76 for *angry_why*. On the contrary, the *brkdhlp* dataset seems easy to tackle by simple boolean rules, as shown by the .747 accuracy figure of the boolean method; in this case RAINBOW underperforms the boolean method by 13 %, while MCSVM virtually delivers the same performance as the boolean method. Moreover, the supervised learning approach delivers a more stable performance across the three datasets, since the reductions in standard deviation with respect to the same best-performing method are very significant, a -28% for RAINBOW and a –10% for MCSVM. These improvements are even more significant once we remember that they are obtained by a method that is much cheaper than the dictionary-based method in terms of expert human resources.

The fact that improvements of this order of magnitude are obtained even with a method, such as the naïve Bayesian technique implemented in RAINBOW, which is known as an average performer in the text categorization literature, bears witness to the superiority of the supervised learning approach to survey coding.

The fact that multiclass SVMs, known top-performers in the machine learning literature, outperform RAINBOW only by a very small margin, is more surprising. A possible explanation to this might be that the terms occurring in the NORC corpora exhibit low statistical mutual dependence, hence approximating the independence conditions under which Bayesian approaches are theoretically optimal.

6. Conclusion and future work

We have shown that automatic coding of responses to open-ended survey questions may be posed as a multiclass text categorization problem, and that text categorization techniques based on supervised learning may significantly outperform dictionary-based techniques, such as those used in Viechnicki (1998) that have been up to now the dominant approach to automated survey coding. Another advantage of the supervised learning approach with respect to the dictionary-based approach, which requires that the text classifiers be hand-crafted (by a knowledge engineer and a social scientist working together), is that the classifiers can be generated automatically from the training data, with substantive savings in terms of expert human resources.

The effectiveness levels that text categorization techniques have achieved in our experiments are far from being perfect, and also from being completely satisfactory. Although the results obtained in our research are promising, we think that more research is needed for the automatic approach to survey coding to clearly supersede the manual approach.

Acknowledgments

The open-ended text used in this work was collected in the General Social Surveys of the National Opinion Research Center (NORC), University of Chicago, and supplied by NORC to the authors. We are grateful to Tom Smith and Jennifer Berktold for providing these texts and for assisting us in their interpretation. We are also grateful to Koby Crammer, Yoram Singer and Andrew McCallum for making the MCSVM and RAINBOW packages available, to Peter Viechnicki for clarifying several points of his experiments, to Henri Avancini for helping with several pre-processing issues.

References

Brank J, Grobelnik M, Milić-Frayling N and Mladenić D (2002) Interaction of feature selection methods and linear classification models. In *Proceedings of the ICML-02 Workshop on Text Learning*, Sydney, AU.

Crammer K and Singer Y (2001) On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, 2, 265-292.

Davis J A and Smith T (1996) *General Social Surveys, 1972-1996: Cumulative Codebook*. National Opinion Research Center, Chicago, US.

Dumais S T, Platt J, Heckerman D and Sahami M (1998) Inductive learning algorithms and representations for text categorization. In Gardarin G, French J C, Pissinou N, Makki K and Bouganim L (Ed.), *Proceedings of CIKM-98, 7th ACM International Conference on Information and Knowledge Management*, 148-155, Bethesda, US. ACM Press, New York, US.

Fuhr N, Hartmann S, Knorz G, Lustig G, Schwanter M and Tzeras K (1991) AIR/X - A rule-based multistage indexing system for large subject fields. In Lichnerowicz A (Ed.), *Proceeedings of RIAO-91, 3rd International Conference "Recherche d'Information Assistee par Ordinateur"*, 606-623, Barcelona, ES. Elsevier Science Publishers, Amsterdam, NL.

Giorgetti D and Sebastiani F (2003) Automating Survey Coding by Multiclass Text Categorization Techniques. *Journal of the American Society for Information Science and Technology*. Forthcoming.

Hsu C W and Lin C J (2002) A simple decomposition method for support vector machines. *Machine Learning*, 46 291-314.

Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In Nédellec C. and Rouveirol C. (Ed.), *Proceedings of ECML-98, 10th European Conference on Machine Learning*, pages 137-142, Chemnitz, DE. Springer Verlag, Heidelberg, DE. Published in the "Lecture Notes in Computer Science" series, number 1398.

Lewis D D (1992) An evaluation of phrasal and clustered representations on a text categorization task. In Belkin N J, Ingwersen P and Pejtersen A M (Ed.), *Proceedings of SIGIR-92, 15th ACM International Conference on Research and Development in Information Retrieval*, 37-50, Kobenhavn, DK. ACM Press, New York, US.

Li H and Yamanishi K (2002) Text classification using ESC-based stochastic decision lists. *Information Processing and Management*, 38(3), 343-361.

Macchia S and Murgia M (2002) Coding of textual responses: Various issues on automated coding and computer assisted coding. In *Proceedings of JADT-02, 6th International Conference on the Statistical Analysis of Textual Data*, 471-482, St-Malo, FR.

McCallum A K and Nigam K (1998) A comparison of event models for naive Bayes text classification. In *Proceedings of the 1st AAAI Workshop on Learning for Text Categorization*, 41-48, Madison, US.

Salton G and Buckley C (1988) Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5), 513-523. Also reprinted in (Sparck Jones and Willett, 1997), 323-328.

Schütze H, Hull D A and Pedersen J O (1995) A comparison of classifiers and document representations for the routing problem. In Fox E A, Ingwersen P, and Fidel R (Ed.), *Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval*, 229–237, Seattle, US. ACM Press, New York, US.

Sebastiani F (2002) Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1-47.

Sparck Jones K and Willett P (Ed.) (1997) *Readings in information retrieval*. Morgan Kaufmann, San Mateo, US.

Taira H and Haruno M (1999) Feature selection in SVM text categorization. In *Proceedings of AAAI-99, 16th Conference of the American Association for Artificial Intelligence*, 480-486, Orlando, US. AAAI Press, Menlo Park, US.

Tzeras K and Hartmann S (1993) Automatic indexing based on Bayesian inference networks. In Korfhage R, Rasmussen E. and Willett P. (Ed.), *Proceedings of SIGIR-93, 16th ACM International Conference on Research and Development in Information Retrieval*, 22-34, Pittsburgh, US. ACM Press, New York, US.

Viechnicki P (1998) A performance evaluation of automatic survey classifiers. In Honavar V & Slutzki, G (Ed.), *Proceedings of ICGI-98, 4th International Colloquium on Grammatical Inference*, 244-256, Ames, US. Springer Verlag, Heidelberg, DE. Published in the "Lecture Notes in Computer Science" series, number 1433.

Yang Y and Liu X (1999) A re-examination of text categorization methods. In Hearst M A Gey F and Tong R (Ed.) (1999). *Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval*, 42-49, Berkeley, US. ACM Press, New York, US.

About the Authors

Daniela Giorgetti is a researcher at the Computational Linguistics Institute of the Italian National Council of Research in Pisa. Her current research interests are focused on the use of text categorization and text mining techniques in linguistics related fields, while her previous background is in computer

graphics, multimedia and virtual reality models, applications, and interfaces. She can be contacted at ILC-CNR, via Moruzzi, 1, 56124 Pisa; tel. +39 050 315 2841; fax +39 050 31 2839; e-mail daniela.giorgetti@ilc.cnr.it.

Irina Prodanof is a senior researcher at the Computational Linguistics Institute of the Italian National Council of Research in Pisa. Her research interests range from man machine interaction to dialogue modelling, information extraction and multilinguism. She can be contacted at ILC-CNR, via Moruzzi, 1, 56124 Pisa; tel. +39 050 315 2866; fax +39 050 315 2839; e-mail irina.prodanof@ilc.cnr.it.

Fabrizio Sebastiani is a researcher at the Computer Science Institute for Information Science and Technologies of the Italian National Council of Research in Pisa. His main current research interests are at the intersection of information retrieval and machine learning, with particular emphasis on text categorization, text mining, and information extraction. He can be contacted at ISTI-CNR, Via Moruzzi, 1, 56124 Pisa, Italy; tel. +39 050 315 2892; fax +39 050 855838; e-mail fabrizio.sebastiani@isti.cnr.it.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Active Prototypes. Modelling Survey Metadata

Stephen Jenkins & Tony Solomonides

Abstract

This paper describes a new and flexible survey metadata model which allows for a high degree of user extensibility and programming. At its foundation is a prototype-based object model. It uses prototypical objects to model both system-supplied and user-generated functions and operations. The paper describes the principles of prototype-based systems and connects with work previously done on prototype-based programming languages. The significant benefits accruing by the application of these techniques will be discussed and explained.

Keywords

Metadata; prototype-based object model; domain-specific programming

1. Background

The very first survey software systems were driven by specialists that were a hybrid between the programmers who wrote the systems they used and the users of their output—the commissioning researchers. Those researchers, in turn, worked for a client organisation or individual. In those early days the role of survey software was largely the analysis of results.

The software role is continuing to change and evolve to routinely encompass more and more of the survey process: certainly to manage the stages of questionnaire design, respondent interviewing, data collection and subsequent analysis; possibly extending to the publication of high-quality summary tables and charts with associated narratives and reports. Furthermore, many of the users of the software nowadays are the individuals that were the clients in the original scenario.

Those users are of varying aptitude and focus. Some require significant help in setting up individual survey projects. Such users would benefit from the use of pre-configured standard components representing individual questions, entire questionnaires or even whole surveys complete with outline summary analyses. Others, given suitable tools, would be quite capable of devising the components for those other users to deploy.

An essential observation in this view is that those developing the components are developers of software but they are not programmers. They want to work (devise components) without leaving the survey domain. They do not want to be faced with a decision that in order to build such-and-such component requires them to be expert in some other, possibly alien, computing domain and its associated development environment.

The ability to devise components without leaving the survey (or other) domain paradigm is fundamental and applicable to many domains. It is the solution to this problem that sits at the crux of this paper.

Within the general goal of devising an environment in which domain-level components may be devised and constructed, there are a number of additional or associated attributes we consider it would be useful to accommodate:

- A desire for users to be able to describe what needs to be done, not how to do it.
- An ability for everything (question, table, calculation, function application) to incorporate a description of its result—the semantics of results is important in determining conclusions to be drawn.

One notable aspect of software development, of whatever kind, is that it is often easier for a user (whether designer or developer) to devise a specific, narrow solution to a problem and then later develop a more generalised solution from that. This applies to the development of software artefacts in the survey domain as much as any other. For example, suppose that a survey included a number of questions requesting the respondent to quantify the amount spent on various goods (such as cereals, vegetables and meat). The user may want to categorise those responses into monetary bands for the purposes of performing some summary analysis. One way to do this is to develop an appropriate banding scheme 'hard-wired' to one of the source questions (for example, the one on cereals), then to copy this banding scheme and make the necessary text expression changes to have it refer to one of the others (for example, the question on meat expenditure).

With a little more work, a suitable user could have generalised the banding scheme to make it applicable to any of the anticipated source variables in such a way that it acts as a function prototype. The foundations of this prototype-based view are discussed next.

2. Object Programming

Object-oriented programming is now firmly established as the paradigm of choice for programming commercial systems and software products. Objects are identified by a correspondence between the objects found in the application domain and are used to model appropriate domain operations. For example, devising and writing a program for manipulating surveys may involve developing software objects to represent tabulations and questions and so on. Each question object would be responsible for managing its configuration and appearance in an interview (which would also be an object). It would also be responsible for performing validation of respondent replies and for giving up the text of labels associated with answer-choice categories when requested to do so by a tabulation object. The behaviour of the program as a whole is then determined largely by the interactions between these internal objects.

The essential characteristics of object-oriented systems are that they exhibit at least the following three properties:

- Encapsulation—A reflection of the desire to keep the definition of all behavioural aspects of an object in one place and a requirement to distinguish between data and functions used internally from those supplied externally to other objects. The intent of encapsulation is to distinguish between **what** an object does (the external view) from **how** it does it (the internal part).
- Inheritance—A mechanism by which objects may share information based on a notion of specialisation and generalisation. For example, an open-response question expecting a numeric response could be seen as a specialisation of an open-response question expecting a free-text response. Arrangements may be made such that the numeric question inherits many of the properties of the free-text question but specialises it by applying an additional constraint on

allowable response. The inheritance relation is sometimes known as the 'is a' relationship—a 'numeric response question' is a 'free-text response question' but with additional features.

- Polymorphism—An observation that objects which are connected by the inheritance relationship are substitutable for one another. Thus if an operation was written to expect a 'free-text response question' object, a 'numeric response question' could be given in its place. This is a natural extension of the 'is a' view of inheritance, above.

Class-Based Object Programming

Languages such as Java and C++ are commonly used to express object-oriented programs although many others (including Visual Basic, C# and Eiffel) are based on the same principles. All of these languages are known as *class-based* languages. Their approach is based upon the identification of common properties of objects and describing them in a defining structure known as a *class*. The software objects themselves which are generated by these definitions are known as *instances* of the class. They are generated by a process of *instantiation*.

Thus a class may be created to act as a definition of a typical question. It may define properties such as the text of the question and the type of response expected. Specific questions would be realised by instantiation from the class. Each would have an assigned text (such as "What was your age last birthday?") and an assigned response type (continuing the example, an integer number perhaps constrained to between 0 and 120).

This class-based object programming methodology requires that the classes be identified and defined before they can be instantiated. That in turn means that there is a clear distinction between a design phase and a 'use' or 'run' phase. The classes are created during the design phase, the objects are created (instantiated) during the run phase.

Prototype-Based Object Programming

Prototype-based (or object-based) object programming languages are distinguished from class-based languages by the lack of explicit classes. Instead, one object is identified as acting as a prototype of many similar but distinct objects. Each of those distinct objects can themselves be used as prototypes for further objects and so on. The term prototype is used in the sense that the prototype is regarded as a reference or example object. The lack of the notion of classes means a simpler conceptual view from a user's perspective. It also enables a system design to be built incrementally and by example rather than by having distinct design (class definition) and system run (object instantiation) phases.

It has been shown that although explicit classes are essentially 'missing', the prototype-based object paradigm is capable of representing everything that the class-based object paradigm can and more. The key benefit that the prototype-based object model brings is that it provides the user with a single environment with which to compose objects. It makes it as easy to specialise an instance from a prototype as it does to generalise a prototype from an instance.

From a system composed initially of suitable basic, or primitive, objects (such as 'a question', 'a table', 'a questionnaire' and so on), users are able to deploy those objects, develop new compound forms and extend the primitive forms as they see fit. For example, if a particular application required that all objects include a "notes" field, that could be incorporated at the appropriate root level. That is not simply a decoration which would apply to that single object. Because the system follows the usual notion of inheritance then such a notes field would be available for every instance or prototype deriving from that root.

3. Modelling Survey Objects Using Prototypes

The previous section took a fairly abstract view of the possibilities afforded by the prototype-object model. In this section we consider some actual applications. Initially these applications are those that relate to survey questionnaires.

Grid or 'Battery' Questions

A grid question is one where a single question, or a composite question, is asked repeatedly for a prescribed number of topics. Abstractly, a grid may be considered to be based upon a prototypical question which represents the 'typical' grid row definition. The questions for each of the aspects may then considered to be instances of that prototype. Any changes made to the prototype would automatically be reflected in each of the aspects using the standard inheritance mechanism. The situation is illustrated in Figure 1.

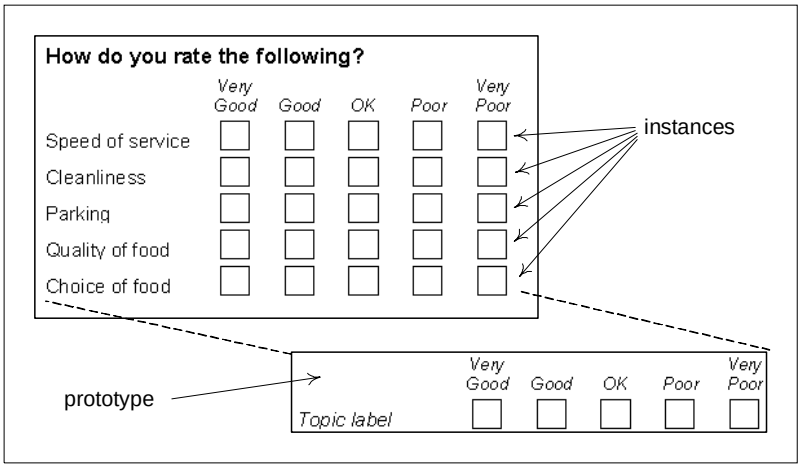


Figure 1 A Grid Question based on a Prototype

Sometimes grid questions are used in circumstances where each aspect iterated over does not share exactly common attributes with its peers. For example, a survey of canned drinks might identify three drinks as being available in "small", "medium" and "large" sizes and one as being available only in "medium" and "large" sizes. In those circumstances, the individual aspects implement their own versions of the prototype. They do so using exactly the same mechanism as would be used to hide inappropriate answers at interview time for a particular respondent. It is just that in this circumstance, the affected answer category (or categories) would be hidden for *all* respondents—that is, the `hide if` property of the relevant categories would be set to the constant value `true`.

Importantly, for this (prototype-based) view to work, the prototype is not of itself a question in the questionnaire but the objects instantiated from it which form the individual aspects *are* questions.

Questionnaires

When researchers devise and design questionnaires they are creating a template for the interviews that may be conducted with it. In the simplest questionnaires, where there are no dynamic elements, the respondent is presented with each of the original elements (questions and messages) of the entire questionnaire. In many instances, especially where interviews are conducted electronically, there will be a number of specialisations made to the original design to target it more precisely to each particular respondent. These may effect specialisation of any of the following:

- Wording: question texts and labels may be modified to include responses to other questions or other relevant data relating to the respondent which is already known.
- Routing: individual questions may be excluded as routing directives take effect.
- Consistency: individual answer categories may be removed as consistency imperatives take effect.
- Derivation: other required derived values will be calculated as appropriate answers become available.
- Rotation: In order to reduce bias effects of presentation orders answer categories may be moved within questions and / or questions may be moved within blocks.

These are specialisations that occur at interview-time. But there may be others required before then. Notably there may be specialisations applied to the wording for language (English, French, German etc.) and for interview mode (paper, Internet, PDA etc.).

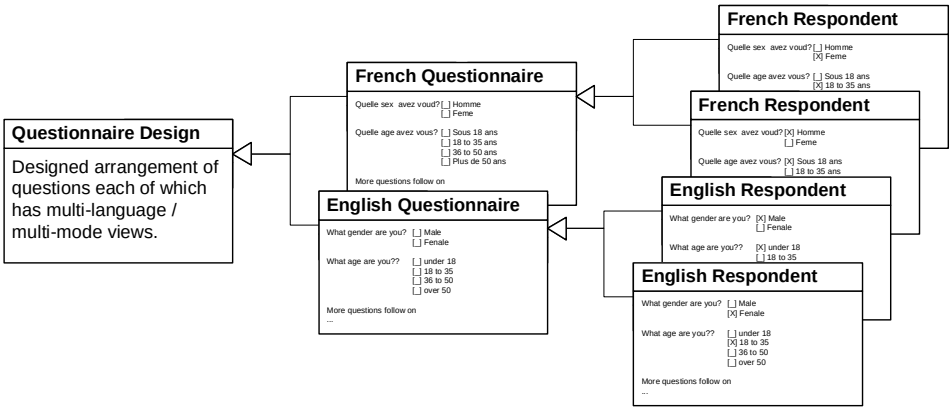


Figure 2 A Questionnaire Design and its Instances

Specialisations for language and interviewing mode are typically required between the composition of the original design and the specialisations introduced for each respondent (See Figure 2). All can be viewed as developing specialisations of the original design questionnaire, initially for language and interviewing mode, and ending at a point where specialisations are unique to each respondent. The original questionnaire is thus acting as a prototype from which the others are ultimately instantiated.

Abstract Questions

The same prototype-instance relationship may also be seen between the abstract notion of, say, a ‘quantity’ question (one expecting a numeric value as a response) acting as a prototype for all ‘quantity’ questions. When an actual question is devised, the question name or number, the phrasing and any required restrictions on the range of acceptable responses, are the specialisations applied to the abstract prototype.

The abstract prototype questions are provided by the survey software application itself. As well as being specialised directly as questions in a questionnaire, they could be specialised to create standard (but possibly parameterised) library questions by a suitably qualified user. The library questions would in turn act as prototypes for actual questions deployed by an end user in their questionnaire design.

These arrangements are illustrated in Figure 3, in which two things that are apparent:

- The use of domain library components has simplified development of the composed questionnaire: All that is required for definition of *Q2* is specification of the appropriate question text. In another circumstance, it is quite possible to conceive of question libraries holding questions that include the complete wording—indeed that may be considered an essential requirement if composed surveys are to be subjected to benchmark comparisons.

- The semantic level of the composed questionnaire is raised: Interpreting the inheritance (specialisation) connection as an 'is a' relationship, *Q1* 'is a' *Numeric Question* and would thus respond to appropriate operations involving numbers (such as addition, multiplication, etc.). On the same principle, the *Q2* 'is a' *Age Question* and would therefore respond correctly to appropriate operations. Such operations might invalidate multiplying one age by another for example.

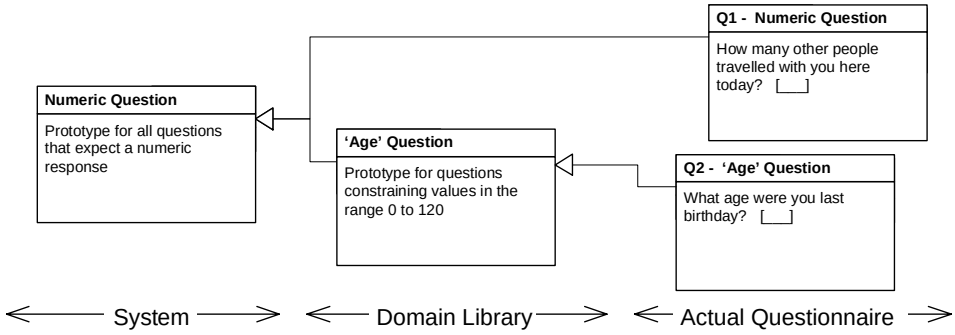


Figure 3 Abstract Question and Instances

Of importance, at this stage, is that we are seeing the beginnings of the notion of programming the system where some users are able to build specialised components for others to deploy. Furthermore, the component-building users are composing their objects in exactly the same (survey) domain environment that their creations are deployed in—and they do not have to become expert in some other programming language or system in order to do so. The next sections further illustrate this point.

4. Survey Variables, Derivations and Functions

At the end of Section 1 it was proposed that a user wanting to develop a banding scheme for a numeric question could develop such a scheme for one numeric variable and then generalise the resultant object so that it was applicable to many such source variables. In this section we take that general example but apply it to analysis of attitudes.

A common requirement (in some survey domains) when analysing attitudinal data is to count each of the individual response categories, calculate a mean score (and sometimes other statistics) by applying a scoring system to each of the categories and additionally summarise the 'overall positive' and 'overall negative' categories. Figure 4 shows a typical example.

The example is fundamentally a crosstabulation of responses to an 'opinion' question forming the rows and an 'age' question forming the columns. At the 'age' question, respondents were asked to nominate the range group their age falls in.

The rows of the table are actually derived by applying a function to an 'opinion' question similar to one of the topics from the grid question shown in Figure 2. The function applied is actually a compound of three functions: a calculation of a mean value, a re-iteration of the original input variable and a summarising (recoding) variable. The result is a new variable object, illustrated in Figure 5, which generates the required result from the question variable to be analysed.

Absolute Respondents	Base	Age				
		Under 18	18-24	25-44	45-60	Over 60
Base	204	59	40	41	44	20
Opinion						
Mean	0.88	1.27	1.30	0.71	0.55	0.00
Very Good	52	24	16	4	4	4
Good	96	27	20	25	20	4
OK	40	8	4	8	16	4
Poor	12	-	-	4	4	4
Very Poor	4	-	-	-	-	4
Positive	148	51	36	29	24	8
Negative	16	-	-	4	4	8

Figure 4 Extended Analysis of Opinion

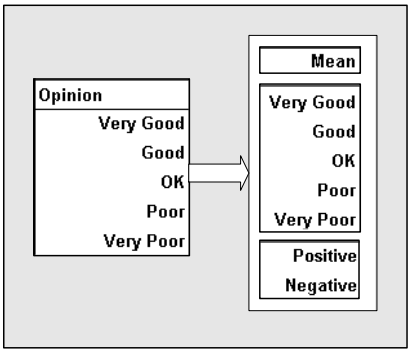


Figure 5 Recoded Variable Object

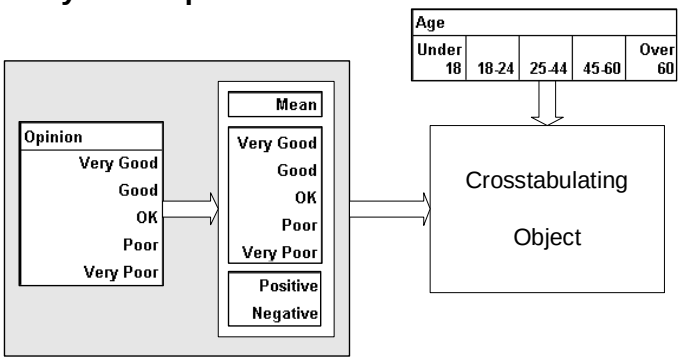


Figure 6—Composing a Crosstabulation

The composed recoded variable object and the 'age' variable are subsequently applied as appropriate arguments to a system-supplied 'crosstabulation' prototype object. This arrangement is shown in Figure 6. The table will be composed as an instance of the crosstabulation prototype object specialised by the variables given as arguments to the row and column parameters. The final table is thus derived by specialisation from the prototype crosstabulating object. The reverse process—generalisation—may be performed on the recoding function. If the source of the recoding function is made a parameter rather than being fixed to the one question variable, then the function may be reapplied to other candidate source variables.

Thus, in the first place, the recoded version of the variable may have been constructed using a question such as "Did you think that the speed of service was …". Generalising that specific result by making the source variable into a parameter means that other appropriate questions may be applied to the recoding function to derive a similar analysis of each. On every occasion that such an application takes place a new recoded variable object will be formed. That new object may be seen as an instance of the parameterised function acting as a prototype, specialised by the particular source (opinion) variable used as an argument.

5. Building and Using Larger Components

The kind of function described in the previous section is a common requirement of many surveys. It may even be the canonical analysis for many. In such circumstances it is never a requirement to analyse the simple, original variable but always a requirement to derive the extended form of analysis.

It would be tedious if every time the user wanted to analyse one of those attitudinal questions, they had to remember to apply the appropriate transformation function. It would be much more convenient

if there was a specific type of table that always, say, treated the variable forming the row axis in that way. Such a table would be seen as a component in exactly the same way that the generic 'crosstabulation' object is viewed. In this section we consider the composition of such a component from both theoretical and practical perspectives.

The Practical Perspective

When the table in Figure 4 was composed, a dataflow path was described starting with the source question, processing through the given recoding function and thence into the table building function as illustrated in Figure 5. The view up to now has been that the argument to the prototype table was the combination of the recoding function applied to the question. That is, the question object and recoding function object were bonded together and supplied the required argument to the tabulation.

But the connection could be formed differently. If the recoding function (alone) is applied to the appropriate axis parameter of the table building function, another table prototype will have been created—one that still has a similar pattern of unknowns. The two alternative arrangements are illustrated in Figure 7.

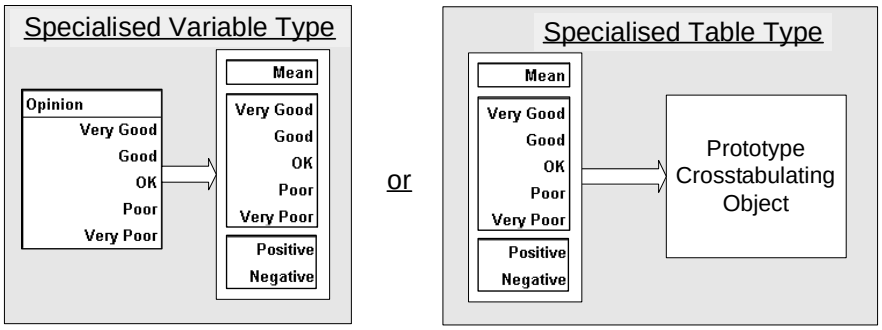


Figure 7 Function Application

By composing a new table prototype in this way a new type of table has been created. The result is an object, formed by specialising the abstract (system-supplied) 'crosstabulation' component, which will be useful in a particular sub-domain. Such an object could be made available as a component in a domain-specific library. Other specialist types of analyses can be similarly composed and made available in the same way. Once again (*c.f. section 3*) the user has the facility to construct components for others to use. They may do so without recourse to any specialist programming; nor are they required to use an external programming language. Importantly, there is no prescribed time at which these components would be created. They could be created as the result of a conscious design decision resulting from a requirement to supply a suitable range of components for a particular sub-domain; or they could emerge as an observation of repeated application of alternative arrangements.

The Theoretical Perspective

The composition of a new table type discussed amounts to a view that the result of composing an object from a variable and appropriate modifying function, then applying the result as the argument to a (tabulating) function arrives at the same result as composing an object from the modifying function as the argument to the tabulating function and then composing another object from that and the variable as the new argument. That is, if ○ is the 'compose' operator:

`(var ○ fun) ○ tab` [1]

is equivalent to:

`var ° (fun ° tab)` [2]

The sub-expression `(fun ° tab)` in `[2]` represents a new type of table in the form of a template object. It acts as a prototype for a succession of table instances produced when different, and appropriate, occurrences of `var` are applied.

Therefore, given a suitable resource of:

var_1, var_2, ..., var_n

and a derived table prototype `newtab`:

`newtab := (fun ° tab)`

then:

var_1 ° newtab
var_2 ° newtab
...
var_n ° newtab
are each individual table instances derived from the `newtab` prototype.

6. Review and Conclusions

The prototype-based object model acts as a powerful foundation for describing survey metadata giving end users the ability to compose components for themselves and others to use. In particular, it confers the following benefits:

- It is *conceptually simple* in principle
- Systems built in this way are both *configurable* and *extensible*

The key benefit of the prototype-based object model is its conceptual simplicity. It does not require knowledge of an abstract layer of object prototypes before full use can be made. Given a supply of suitable initial prototypes (amongst which would be, for example, survey variables and summary analysis tables), users are able to devise specific instances. They are thus able to work with such a system right from the start in much the same way that they would with any other system.

Given a pool of suitable prototypes, actual instance objects may be derived from them. Much of the time this process is a completely normal one, no different from creating a new question or specifying a new summary analysis table in a conventional system. In addition though, further prototypes may be derived by augmenting the initial pool. Users are able to configure the entire system towards specific sub-domains without having to resort to some other underlying programming language or development system.

New prototypes may be simply developed from specific instances: the configuration given to a particular object design often comes from a perspective of actual usage. If a sub-domain expert is able to develop a functional or operational object for a specific task then they, or another user, would be able to convert that specific instance into a more generalised solution.

The ability to modify and evolve objects reduces the need for prior or formal classification and allows for an iterative and experimental development style. As they increase in knowledge and confidence, or simply as their requirements change, some users would be in a position to generate parameterised objects with a view to including them into appropriate domain-specific libraries.

References and Further Reading

Abadi M and Cardelli L A Theory of Objects, Springer Verlag 1996.

Craig I The Interpretation of Object-Oriented Programming Languages, Springer Verlag 2002.

Czarnecki K & Eisenecker U.W Generative Programming, Addison-Wesley 2000.

Finkel R.A Advanced Programming Language Design, Addison-Wesley 1996.

Foote B Objects Reflection, and Open Languages, Workshop on Reflection and Metalevel Architectures Tokyo, Japan. IMSA 1992.

Gamma E, Helm R, Johnson R & Vlissides J Design Patterns, Addison-Wesley 1994.

Lieberman M Using Prototypical Objects to Implement Shared Behaviour in Object Oriented Systems, In Proceedings of First ACM Conference on Object-Oriented Programming Systems, Languages and Applications, Portland, OR, September 1986.

Noble J, Taivalsaari A & Moore I Prototype Based Programming: Concepts, Languages and Applications, Springer Verlag 1999.

Ungar D & Smith R.B SELF: The Power of Simplicity, In Lisp and Symbolic Computation: An International Journal, 4, 3, 1991, Kluwer Academic Publishers 1991.

About the Authors

Dr Stephen Jenkins is Technical Director of Mercator Research Group Ltd. He has a Ph.D. on the subject of metadata models and domain-specific languages for the survey application domain. He is the architect of the snap survey software system and is responsible for its direction and development. He is one of the founding members of the triple-s group and has contributed to the Metanet project and to meetings of the Object Management Group on the subject of survey metadata. He can be contacted at Mercator Research Group Ltd., Mead Court, Thornbury, Bristol BS35 3UW, England; or by telephone on +44 (0) 1454 280800; or by email at `sjenkins@mercatorResearch.com`.

Tony Solomonides is a Reader in Computer Science in the Centre for Complex Cooperative Systems at UWE and a CERN Associate. He has researched scientific databases and applications since the mid-1990's, having also worked in medical informatics, especially the management of chronic diseases, before then. He has participated in the development and application of the CRISTAL database to track the construction of a Large Hadron Collider experiment at CERN. He has also written on the application of 'description-driven' methods and systems to the medical field, notably on large scale Electronic Patient Records, and on the tracking of antibiotic resistant bacteria in the community. He has also published extensively in the field of data collection through social survey instruments. He can be reached by telephone on +44 (0) 117 344 3150 or by email at tony.solomonides@uwe.ac.uk.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Write Global, Run Local. Software Design Issues For Global Data Collection

Paul Quinn & Leonid Sokolovsky

Abstract

The Market Research Industry has changed very substantially in the last 10 years, with a few very large multi-national corporations now commanding a significant share of the market. This change, driven by a combination of significant merger and acquisition activity and organic growth, has resulted in a complex set of problems for these businesses when conducting international and mixed-mode research on a global scale. Typically five or more separate and different data collection systems may be required in order to complete a given project. This may mean that five or more different CATI or CAWI scripts have to be written – functionally equivalent but logically different, yielding five or more different data sets in different physical locations. This gives rise to considerable management complexity, is error prone and time consuming, and often highly complex data manipulation is required in order to normalise the data for subsequent downstream processing and analysis.

What sort of system could be built to overcome these challenges, using contemporary technology and architectural designs?

This paper discusses the design considerations for such a system, as encapsulated in a software development project that has resulted in a new mixed-mode interviewing system called Bellview FUSION.

Keywords

Mixed-mode research; declining response rates; automation scripting; centralised control; distributed architecture; international research

1. Introduction

This paper introduces the many ideas that have gone into the creation of a new sort of data collection system, designed for the increasingly complex, time pressured and global nature of modern market research. This is an industry characterised by rapid consolidation and globalisation, which is somewhat conservative by nature when it comes to technology. It is also an industry that faces continuing and growing shortages in raw material – the market research respondent. Studies indicate a continued decline in response rates over the last 20 years and this paper postulates that technology may have a part to play in at least stemming the flow.

The paper draws on the 17 man-years so far placed in the creation of a new software system, Bellview FUSION, which attempts to address the following key issues.

The requirement for a centrally controlled system

Although data collection sites, such as CATI centres themselves may be geographically scattered, and some degree of local autonomy required, it is also a requirement that the system be centrally monitored and controlled. The architecture must allow for both a fully centralised system, with minimal footprint at remote centres, and one where there are degrees of de-centralisation.

The requirement for a single data repository

All of the data should eventually reside, in a common format, in a single, large, resilient data repository. This eliminates the need to reformat data at the consolidation stage and allows near real-time reporting on results. Although data may be collected in a large number of distributed sites, data should be 'trickled' back to the centre in a common format.

The requirement for customisation and automation

The system should be open, with documented interfaces, and be both programmable and extensible. This will deliver a high degree of flexibility and the ability to integrate the data collection system into the workflow and financial systems of host organisations. The adoption of common standards and an open framework is to be encouraged, so that systems from multiple vendors can be more easily integrated into a seamless operation.

The requirement for mixed-mode data capture

1. Data collected via the Web (CAWI), phone (CATI), paper (PAPI), face-to-face (CAPI) and via Interactive Voice Response (IVR) should all reside in a common format in the data repository.
2. A single multi-mode survey script should drive the system.
3. In an effort to offer more choice to respondents and to increase response rates, they must be allowed to choose their preferred mode of interview, and must be allowed to switch between modes as many times as required. Naturally, the appropriate mode must also be able to be imposed.
4. Considerable meta- and para-data must be maintained by the system in order to unravel the baseline collected data. It is important to capture the mode in which a given answer was collected, as the methodological implications of multi-mode interviewing are still not well understood.
5. Considerable effort must be expended to take full advantage of the unique characteristics of each mode. We must eliminate the 'CATI on the Web' approach which characterised the early days of Web interviewing.

A tightly coupled/loosely coupled system

1. Where robust WAN/LAN links are in place, these should be exploited.
2. However, if a 'local' site loses contact with the centre, it must cache sufficient information to allow it to continue to function in a disconnected state, with a re-synchronisation occurring when the network is re-established.
3. The employment of a distributed workforce must be catered for – it is increasingly common for CATI 'centres' to be de-centralised.
4. Predictive/Auto-diallers at CATI centres must be available to local and distributed agents.

2. Globalisation

In the 2001 Annual Study of the Market Research Industry, ESOMAR estimates that the total Market Research Market was worth some $15.9 billion in 2001. Table 1 below, which lists the top 25 Global

Research Organisations by turnover, shows that these companies between them represent some 61.28% of the total revenues.

This globalisation is in part driven by the needs of the end-users of research data themselves, often researching global brands and requiring execution and common standards, research products and processes around the world.

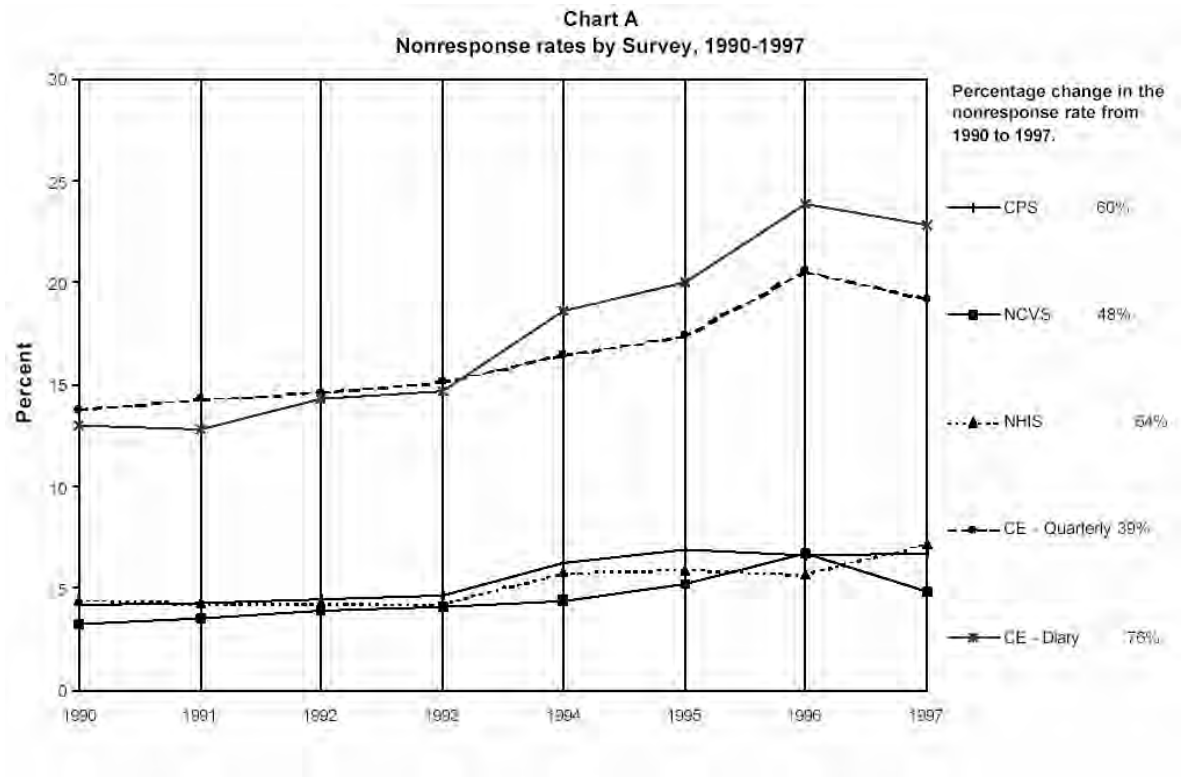
The process of acquisition naturally brings with it the notion of integration, at a great many levels. IT is one of those levels and, like most of the other integration issues, is far from trivial. Aligning a number of geographically dispersed, different CATI systems, for example, whilst at the same time delivering a high level of service to clients requires careful planning and of course a system capable of meeting the changing needs of the organisation.

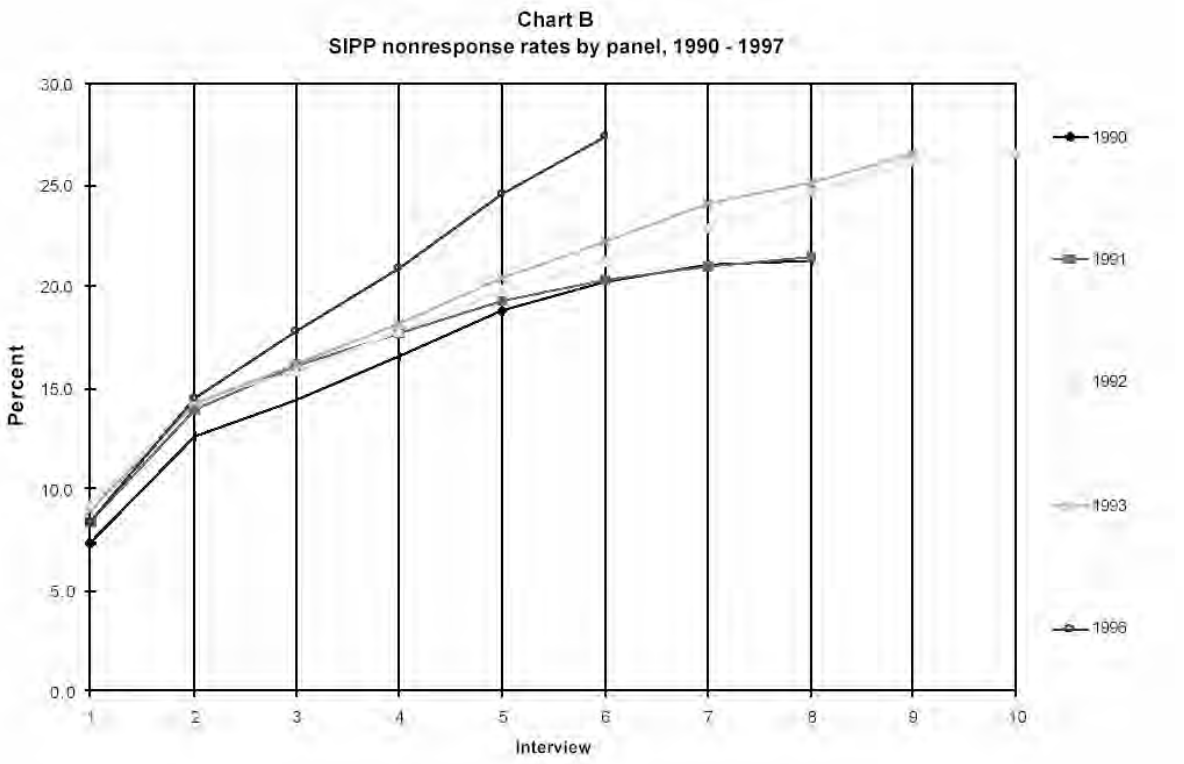
Table 1: Honomichl (2001), Global 25 annual report (pre-dates TNS acquisition of NFO Worldgroup)

2001 Top 25 Global Research Organisations					
2001	Organisation	Global research revenues (US$ millions)	2001	Organisation	Global research revenues (US$ millions)
1	VNU N.V.	2400	14	Video Research Ltd.	162.2
2	IMS Health	1171	15	ORC	133.6
3	WPP plc.	1006.9	16	JD Power	128
4	Taylor Nelson Sofres	813.2	17	Intage	108.5
5	IRI	555.9	18	NPD	101.7
6	GfK Group	479.6	19	Jupiter MM	85.8
7	NFO Worldgroup	452.9	20	Dentsu	78.4
8	Ipsos Group SA	429.9	21	Harris Interactive	75.4
9	NOP World	324.7	22	Abt Associates	62.8
10	Westat	285.8	23	Sample Institute	59.2
11	Synovate	266.5	24	IBOPE	56.1
12	Arbitron	227.5	25	MORPACE Int. Inc.	48.3
13	Maritz	181.7	25	Nikkei Research Inc	48.3

3. Declining Response Rates

It is an inescapable fact that response rates have been and continue to decline. Market Research is seen as less and less relevant to the younger generations. From a study by Atrostic and Burt (1999) the charts below show marked declines in a number of US Federal Government studies, but these patterns are seen in all mature markets and there is anecdotal evidence that the trends have continued to date.





Atrostic and Burt (1999) note: *"During the 1990s, annual nonresponse rates rose markedly (between 40 and 80 percent) in five large, continuing household surveys conducted by the Census Bureau. Chart A shows the annual nonresponse2 rates for the CPS, the CE Diary and Quarterly Surveys, the NHIS, and the NCVS. The percentage changes in rates between 1990 and 1997 for each survey are shown in the legends to the right of the charts."*

A description of these US Federal Government studies is to be found in Appendix B.

"Chart B shows nonresponse rates over the same period for another large, continuing survey conducted by the Census Bureau, the Survey of Income and Program Participation (SIPP). Rates are shown separately for its 1990 through 1993 panels and for its 1996 panel (no panels were introduced in 1994, 1995, or 1997). Nonresponse rates for the initial interview (or wave) of each panel rose from 1990 through 1993. Nonresponse rates for all SIPP panels typically rise quickly with successive interviews." The authors postulate that technology may have a role to play in helping to arrest the general decline in response rates to Market Research surveys over the last 20 years.

1. Mixed-mode interviewing enables respondents to participate in a way and at a time that suits them.
2. Ineligibility for a given study may mean that the sample can be re-used, in a system where all sample, no matter what the mode, is managed as a single entity.
3. Detaching the 'visual' layer from the 'data layer' in a CAWI environment, allows much richer use of the medium, leading to more engaging and visually appealing questionnaire instruments. For example, in asking "Which US states have you personally visited in the last 5 years?", the traditional CATI approach would be to have a multiple-choice answer list with 50 items, and a 'none of these' and 'don't remember' tagged on as single response categories. It would still be common for this to be translated, roughly unchanged, into a web environment. But an alternative approach, much more in keeping with the everyday web experience nowadays, would be to display a map and allow the respondent to click on each state visited, which might then change colour. Behind the scenes, of course, is our old friend the multiple-response variable.

Examining the differences achievable in measured response rate by using mixed-mode surveys, Dillman et al (2001) conclude: *"Results from this analysis leave little doubt that a mixed-mode strategy of following a complete data collection strategy by one mode, with a short pause, followed by an attempt to collect data by another mode, can increase response rates dramatically."*

4. Old and new approaches

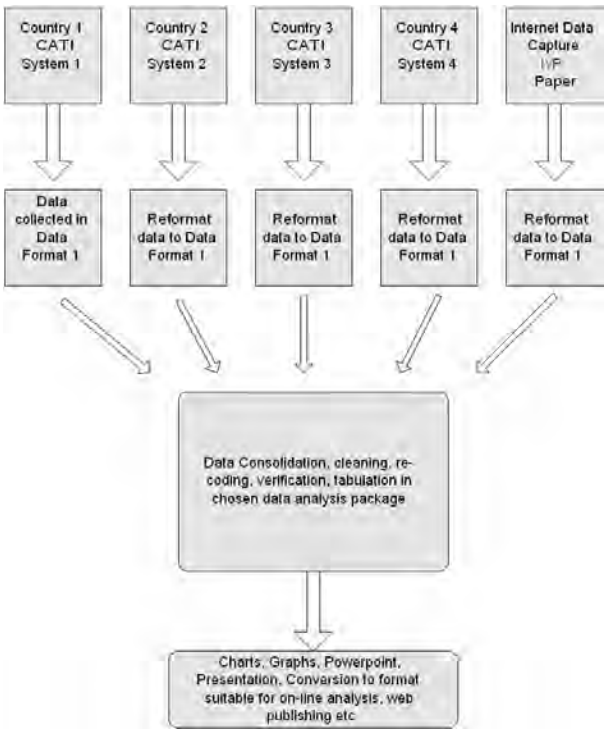


Figure 1: High-Level Traditional Workflow

Figure 1 illustrates the traditional workflow in an organisation that has acquired a number of disparate systems. In delivering an international study, whether it be single-mode or multi-mode data collection, considerable time, effort and expense is incurred in the setup and execution of the job. This is because, effectively, the job has to be set up N times, independently, where N represents the number of systems that have to be used to ensure the required international coverage. It has been estimated that as much as 40% of the cost of data collection could be eliminated by moving to a more streamlined approach such as that indicated below in Figure 2.

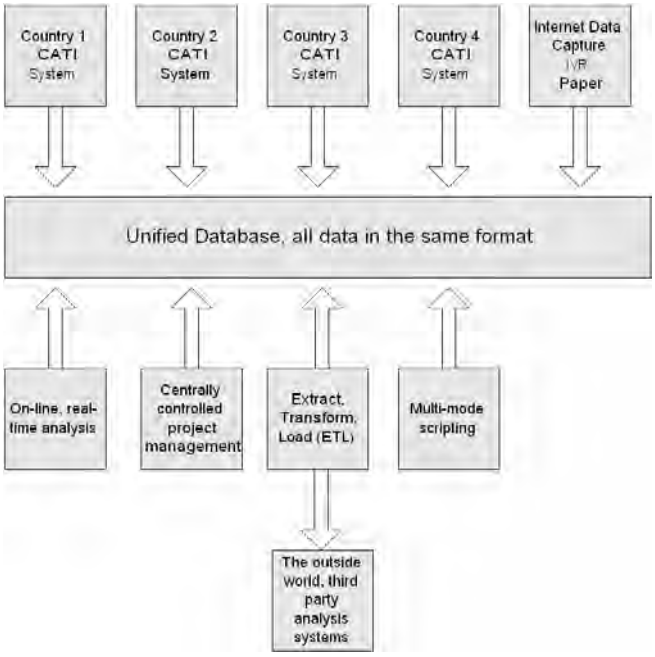


Figure 2: The New Workflow

In this new workflow, the key point is essentially that all of the data is collected in a single place, in the same format, no matter what the mode of collection. This eliminates the data re-formatting process that typically follows data collection in the old workflow, and also greatly reduces the chance for error.

Note that this *does not* imply that the whole system is necessarily centralised, however. Merely that the ultimate resting place for the data is at the centre – it may be that the data resides temporarily on intermediate systems which periodically replicate the data to a central repository/warehouse.

5. Technology

In Bellview FUSION, which is typical of this new generation of mixed-mode data collection systems, data resides within a single large resilient database – Microsoft SQL Server 2000, Enterprise Edition. To enable data population from numerous and very diverse sources there is a complete Component Object Model (COM) model for the programmatic access to the system; data, meta-data and para-data can be accessed, appended to and removed via programmatic automation routines. The COM is a software architecture that allows applications to be built from binary software components. COM is the underlying architecture that forms the foundation for higher-level software services, like those provided by Object Linking and Embedding (OLE). OLE services span various aspects of commonly needed system functionality, including compound documents, custom controls, interapplication scripting, data transfer, and other software interactions.

Although COM is now being superseded by .net, it is still in widespread use; naturally we expect much of the currently COM based software to achieve transition to .NET over the coming years.

6. A centrally controlled system

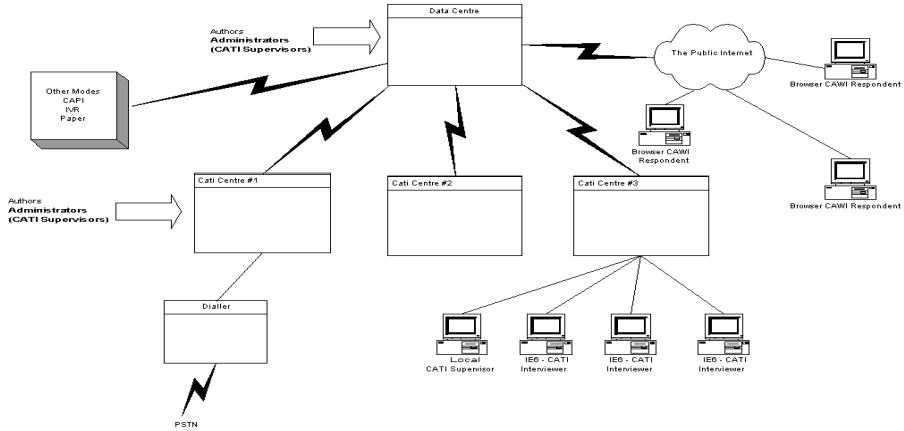


Figure 3: A Centrally Controlled System

The concept of centralised control is the key to a successful globally distributed system. Although there is not necessarily the need for high speed networking and an expensive infrastructure between all locations, it will be necessary to have interconnectivity between the central site and each of the remote locations. These remote sites will periodically and regularly connect to the central site to exchange information.

A centrally controlled system eliminates many of the problems associated with running global projects; the questionnaire is ultimately managed from a single location with sample and user management being collectively controlled. At the same time, however, local sites do need to operate with a certain level of autonomy. Configurable interview record pool sizes would be checked out to a particular remote site to allow this site to operate independently of the central site for a certain period of time. Although ultimately control is centralised, individual sites do need the ability to control call delivery to users and may require localised quota targets to be set.

Careful management of quota targets from a common location reduces the opportunity for over achieving quota targets. Quota checking requires an online connection to the central server at the moment the quota check is called. However, to help eliminate problems when sites are offline, details regarding filled quota cells are pushed to the remote sites as and when they become full, enabling quota failing even when offline. Also, respondents that have agreed to participate in a survey but are quota failed could automatically be reassigned to other projects that require users from within their particular demographic breakdown.

7. The user model

The user model should be based on the Windows model with the concept of local and global users and groups. Security within the system, which is critical, must support the concept of delegated authorities; i.e. based on login profiles such that users are attributed certain views and certain rights.

This hierarchical organisation of users, projects, systems and other resources provides a single point of management, which dramatically reduces redundant, non value-adding tasks and increases accuracy by allowing the management of groups of objects rather than individual entities.

8. A single data repository

There are many benefits offered by a single data repository. However, to achieve a successful operating environment it must be *totally* resilient. Ideally the SQL Server is clustered and carefully managed, especially in an enterprise environment. It is imperative that the central server(s) are very well specified and should reside in a secure and robust physical environment, such as that provided by many of the hosting companies that arose during the dot com era. The smooth running of the central repository is critical to the overall success of the system, and it is likely that considerable Data Base Administrator (DBA) resource may be required in order to achieve high availability and performance.

9. Customisation and automation

Nowadays it is almost impossible to find an out-of-the-box solution to meet an organisation's needs; instead it should be possible to pick and choose best of breed components. For this to succeed the products chosen must provide open interfaces to facilitate integration with existing and new systems.

Products like Bellview FUSION incorporate automation scripting based on well known languages, such as JavaScript and VBScript, with the built-in ability to check/debug syntax. These languages allow the ultimate in flexibility and ensure that a totally customisable solution is delivered. There are many opportunities for automation scripting within the system with entry points based on a series of triggers such as: user login and logout operations, survey open and closing actions, quota cell dynamics, periodical or scheduled routines, data import and export. They also allow the necessary and integrated interaction with systems such as payroll or administrative systems. There are many examples where automation routines can be used to enhance the operation of a system such as the periodic checking and top-up of interview records based on a dynamic total pool size.

10. Mixed-Mode data capture

The ability to perform mixed-mode data collection is the key to a successful, globally deployable solution. Ultimately, all methodologies should be supported and there must be a common format for this data to be described, residing in a single location.

As discussed earlier it is highly desirable to strive for an increase in response rates; consequently it is beneficial to allow respondents to choose their preferred mode of interview. Respondents must also be allowed to switch between modes as many times as needed and at any point in the survey as required. There must be built-in support for the notification of access to CAWI surveys, typically via e-mail and there should be native support to schedule follow up appointments in other collection modes, for instance in CATI or CAPI, if CAWI surveys are not completed within a defined period.

Considerable meta- and para-data must be maintained by the system in order to unravel the collected baseline data. This valuable meta- and para-data must be stored at the question level, must be easily accessible in real time and should be available when exported to particular formats.

11. A tightly coupled/loosely coupled system

Message Queuing and Distributed Systems

Microsoft Message Queuing (MSMQ) technology enables applications running at different times to communicate across heterogeneous networks and systems that may be temporarily offline. Applications send messages to queues and read messages from queues. MSMQ provides guaranteed message delivery, efficient routing, security, and priority-based messaging. It can be used to implement solutions for both asynchronous and synchronous messaging scenarios.

Many distributed applications need the ability to handle delays between a request and a response. This is because all the steps in a distributed application process may not need to be, or cannot be, completed at one time. MSMQ allows applications to use components that communicate with one another using queued messages. Like e-mail messages that sit in an inbox, messages can exist on dissimilar systems that may not even be directly connected to each another.

With MSMQ, applications can be written that do not require immediate responses from either clients or servers, which gives the flexibility needed to handle routine pauses within business processes. Using this service also helps applications to be created that are more readily available and more scalable, and that distribute well over multiple, even unreliable, networks while maintaining the integrity of the data.

Messaging with MSMQ is a flexible, reliable approach to communication - one that's appropriate for many kinds of applications including a distributed CATI/CAWI system such as Bellview FUSION. The system architecture takes care of the queuing processes even if the client and server are not running at the same time. Examples of systems where MSMQ can be used include:

- Mission-critical financial services, such as electronic commerce
- Embedded and hand-held applications, for example, underlying communications to and from embedded devices that route baggage through airports by means of an automatic baggage system
- Sales, for example, sales automation applications for travelling sales representatives
- Workflow. MSMQ makes it possible to create a workflow that updates each system. A typical design pattern is to implement an agent to interact with each system. Using a workflow-agent architecture also minimizes the impact of changes in one system on the other systems. With MSMQ, the loose coupling between systems makes upgrading individual systems simpler.

MSMQ implements asynchronous communications by enabling applications to send messages to, and receive messages from, message queues. These applications may be running on the same computer or on separate computers connected by a network. When an application receives a request message, it processes the request by reading the contents of the message and acting accordingly. If required, the receiving application can send a response message back to the original requester. While in transit between senders and receivers, MSMQ keeps messages in holding areas, or queues. MSMQ queues protect messages from being lost in transit and provide a place for receivers to look for messages when the receivers are ready to receive them.

Applications can use MSMQ to send and process messages regardless of whether the receiving application is running or reachable over the network. This might be because of network problems or planned occurrences, such as mobile users who only connect to the network occasionally. Applications can use transactional messaging in MSMQ, and in this delivery mode, MSMQ makes sure that messages are delivered exactly on time, and in the order in which they were sent.

12. The future

The diagram below (Figure 4) represents a highly idealised, future Market Research 'engine room'. Each component is large, complex and highly sophisticated and the realisation of such a system will depend critically upon open standards and the integration of best-of-breed solutions from a number of vendors. The authors do not believe that such sophistication can be delivered in a meaningful way by a single 'integrated' system. Industry initiatives, such as the SPSS MR Dimensions Data Model, triple-S, TabsML and AskML, XML in general and technologies such as .net from Microsoft (representing Web Services) will have key roles to play going forwards.

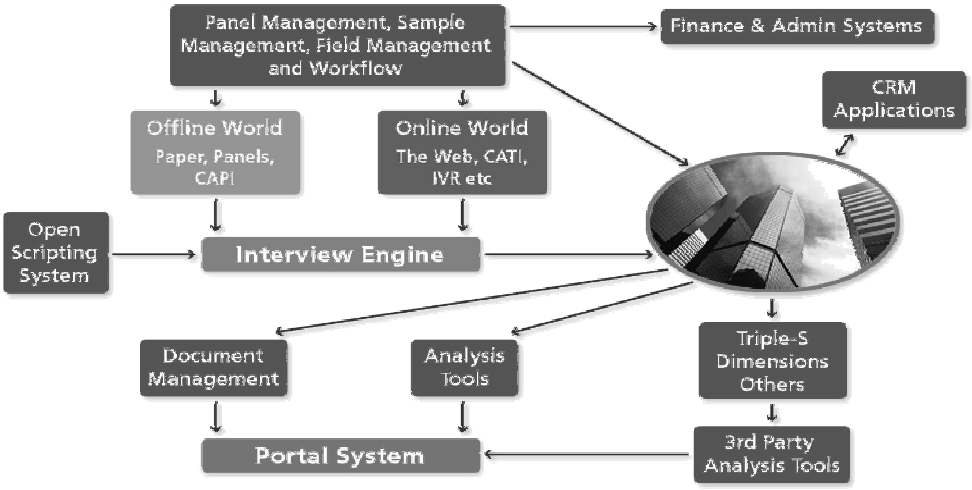


Figure 4: Future Architecture

13. Conclusions and Comments

It is our belief that the ideas presented in this paper, when delivered as working systems to the Industry will help to solve two of the greatest problems currently faced.

1. Declining response rates.
2. The time, cost and complexity associated with global data collection.

We contend, however, that delivery of such systems depends greatly on the emergence of standards and open systems in Market Research, allowing the integration of best-of-breed products to deliver the sorts of functionality implied by Figure 4 above.

Many of the components do exist now, but they suffer from a proprietary nature and hence cannot co-exist in harmony with other elements provided by alternative vendors, at least not in a cost efficient and maintainable manner.

Key vendors in the Market Research software market are making significant steps to improve the situation and we expect to see substantial progress made over the next couple of years.

Appendix A

CE Study. To provide a current and continuous series of data on consumer expenditures and other related characteristics for use in determining the need to revise the Consumer Price Index and for use in family expenditure studies and other analyses.

CPS Study. To provide estimates of employment, unemployment, and other characteristics of the general labour force. In addition to the labour force data, the CPS collects annual data on work experience, income, and migration from the March Supplement, and on school enrolment from the October Supplement. Other supplements, are conducted biennially or intermittently.

NHIS Study. To provide information on a continuing basis about the amount and distribution of illness, its effects in terms of disability and chronic impairments, and the kind of health services people receive. One or more sets of supplemental questions are added each year to gather information on additional major current health issues.

NCVS Study. To provide information on crime victimisation from a general population sample. Data are gathered on types and incidence of crime; monetary losses and physical injuries due to crime; characteristics of the victims; and, where appropriate, characteristics of the perpetrator.

SIPP Study. To collect source and amount of income, labour force information, program participation and eligibility data, and general demographic characteristics to measure the effectiveness of existing federal, state, and local programs; to estimate future costs and coverage for government programs, such as food stamps; and to provide improved statistics on the distribution of income in the country.

Appendix B

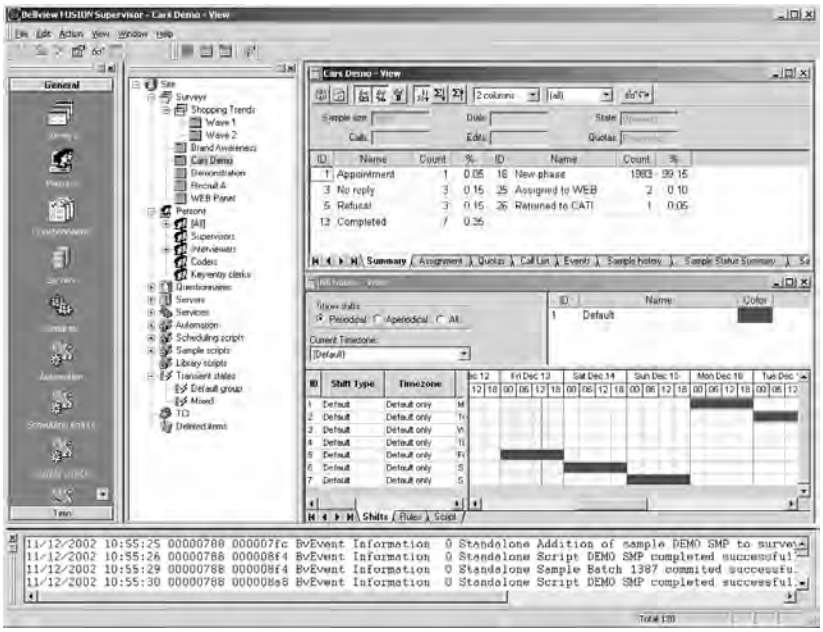


Figure 5: A modern CATI supervisor interface, which supports familiar designs – favourites, multiple windows, drop downs, menus, and customisation.

References

Atrostic B K and Burt G (1999) Household Nonresponse: What We Have Learned and a Framework for Further Work. In Statistical Policy Working Paper 28, Statistical Policy Office, Office of Management and Budget, United States Federal Committee on Statistical Methodology

Dillman D A, Phelps G, Tortora R, Swift K, Kohrell J & Berck J (2001) Response Rate Measurement Differences in Mixed Mode Surveys Using Mail, Telephone, Interactive Voice Response and the Internet, *AAPOR Annual Conference*, Montreal, Canada

Microsoft MSMQ Reference Material
http://www.microsoft.com/windows2000/technologies/communications/msmq/msmq_qa.asp

Samuels J (2001) ESOMAR Annual Study of the Market Research Industry 2001

Honomichl J (2001) Global 25 Annual Report, *Marketing News*, American Marketing Association

About the Authors

Paul Quinn is an associate director at Pulse Train Ltd. He is responsible for the product management of Bellview FUSION and is closely involved in Pulse Train's development strategies. He can be contacted at Pulse Train Ltd., Enigma House, 30 Alan Turing Road, Research Park, Guildford, Surrey, GU2 7AA; tel. 01483 300100; fax 01483 302194; e-mail paul.quinn@pulsetrain.com

Leonid Sokolovsky is Technical director at Pulse Train Ltd. He is responsible for coordinating all Pulse Train's development strategies and is the primary liaison with Pulse Train's innovation centre, Ulter Systems Inc. in Moscow, Russia. He can be contacted at Pulse Train Ltd., Enigma House, 30 Alan Turing Road, Research Park, Guildford, Surrey, GU2 7AA; tel. 01483 300100; fax 01483 302194; e-mail leonid.sokolovsky@pulsetrain.com

SURVEY AND STATISTICAL COMPUTING IV.
Edited by R. Banks *et al*.

Supercollect. A Data Capture System Using an Electronic Network

Hans Stol

Abstract

The completion of questionnaires or request forms, the contents of which must be derived from one or more computerized company records, is not possible in a simple manner. The data of course can be retrieved and entered by hand, although if the data must be provided periodically, this causes a big burden for the companies, being obliged to answer the questions.

One of the policies of the EU countries is to minimise this administrative burden for the enterprises. This is why a few projects were launched in the R&D framework program of the EC, aiming at lowering the administrative burden for the enterprises in filling the questionnaires for the National Statistical Institutes. In the TELER project[1] as well as in the Datamed project[2], the use of EDI in the data collection process of business surveys has been experienced, to solve the problem of the administrative burden. In both projects, electronic questionnaires have been developed to verify the effect on the administrative burden compared with the filling of the questionnaire manually. During the TELER project the Internet technology was not a real option for the data communication part of the data capture process. In the Datamed project a part of the trials has been very successfully carried out by a WEB survey.

Keywords

Electronic data capture, Electronic questionnaire, business surveys, household surveys, XML, Internet

1. Introduction

This paper addresses the problem of the heavy workload of the enterprises that is generated by administrative reporting to fulfil legal obligations. Most often, the enterprises receive paper questionnaires asking for data already stored in their computerised company records, for example numbers of products sold, turnover and number of employees. In the best case, the enterprises can

[1] TELER stands for: Telematics in Enterprise Reporting. This was one of the projects in the 4th framework R&D programme of the European Commission. The main objective of the project was to show how the use of ICT means could lower the administrative burden for the respondents (enterprises) caused by the questionnaires to be filled for the National Statistical Institutes. The following countries participated in this project: Finland, Sweden, Germany, the Netherlands, France, Greece, Italy, Portugal, Spain and Slovenia.

[2] Datamed is an acronym of: DATA-capture in MEDiterrean countries. This was also a project in the 4 framework R&D programme of the EC. National Statistical Institutes of 3 countries were involved. Italy, Portugal and Greece. The DATAMED project investigated the possibilities to use ICT means in getting the data from very small enterprises.

copy the data stored in their own systems one to one towards the paper questionnaire. After filling the questionnaire, it is sent to the data collector, where the data are entered in the computerised system of the data collector.

The use of ICT in the data capture process is the first step in getting a more efficient data collection process. This holds not only for the data collector, being for example National Statistical Institutes (NSI's), other Governmental Institutions or Market Research Companies, but also for the respondent. However, when electronic questionnaires are applied as a "copy" of the paper questionnaire, the enterprises will still not be satisfied. Differences between the questions and the meaning in the data of the different company records as well as differences in storage formats and data structures cause additional problems in the electronic data capture.

Especially as far as enterprises are concerned, the level of standardisation in their bookkeeping systems has to be considered, when we are looking for an efficient data capture implementation, using ICT[1]. In this context, we mean by standardisation: the standardisation of the metadata at the respondents themselves and the standardisation of the metadata used by the data collectors (harmonisation of the meaning of the questions asked for[2].)

The results of the TELER-project showed that the enterprises have automated information systems in which data are stored from which the variables required by the NSI's can be derived[3]. Depending on the country and the kind of variables asked for, the information systems at the enterprises are structured more or less in the same way.

The Internet provided benefits for all categories of respondents:

- For the enterprises as a standard for communications via an electronical network.
- For the VSE's and the households as a mean to use ICT without the obligation of having an in depth knowledge and/or experience in the use of automated information systems.

Examples of the ICT means used in the data capturing process are:

- Electronic questionnaires: depending on the media used (WWW (CAWI), laptop used by interviewer (CAPI), computer used by telephonic interviews (CATI), computer used by the respondent (CASI)).
- Paper questionnaires used in various ways depending on the media used (paper combined with data entry (CADI), paper combined with fax, paper combined with Optical Character Reading).
- Applying direct connection between information systems at the respondent and the NSI, using EDI or e-business-like information systems.

[1] One of the conclusions of the TELER project was that the choices of the means to be used in the data capture process at the enterprises depends on the standardisation of the chart of accounts. These conclusions can be found in the final report of the TELER project.

[2] Metadata play an important role in the statistical process. Especially when that process is or will be highly automated. As of the introduction of the computer in the statistical process the significance of the metadata is increasing. There was no need to make the metadata explicit, when the metadata were in only the mind of the statisticians. Today the use of a correct set of metadata is necessary to design and implement an automated data capture system that serves the data collector and the respondent. An overview of the metadata needed on the various levels (pragmatical, semantical, syntax and medium) can be found in Stol [8]

[3] In the TELER project an investigation has been carried out in ten countries on the availability of the automated information systems. A detailed report on this issue was the result of the first phase of the project: The user requirements. (TELER deliverable 1)

In the matrix below some advantages of the use of ICT for each kind of participant are given.

Table 1. Some effects of the use of ICT in the data capture process

	NSI's	Respondent
Enterprises	No data entry, less control and correction.	Lowest burden: direct extraction from automated information system.
VSE's	Direct control and correction in data entry process. Using WWW, or WEB TV (CASI) by VSE: no data entry, no control and correction.	Lowest burden: via intermediary, fax, WWW
Households	Direct control and correction in data entry process. Using WWW by household: no data entry, no control and correction.	Lowest burden: via intermediary (existing registers and archives), WEB TV, CAPI, CASI

So far, the solutions provided by projects as TELER and Datamed offered an insight into the direction towards concepts by which the administrative burden of the enterprises and the quality of the survey data can be improved. Besides that, the usages of numerous Internet based data capture systems (IDCS) showed the way to use Internet as part of the data capture system.

In this article, we will further discuss the advantages of the use of an integrated data capture system using the Internet, taken into account the lessons learned from the TELER and Datamed project. The main lessons are that an IDCS should provide the following facilities:

1. The enterprises must have the possibility to fill in the questionnaire ***off-line***.
2. The enterprises need a possibility to make an ***automated link*** between the company records and the questions.
3. The automated link should provide the possibility to ***calculate the results*** to be filled in the questionnaire from one or more data items in the computerised records.
4. The automated ***link*** between the company records and the questions should be ***maintained*** over time.
5. Before sending the questionnaire, via Internet, to the data collector the answers should be ***validated*** to prevent the need to ask questions afterwards by the data collector.

2. Requirements of a data capture system using Internet

In this section I shortly describe the requirements for a data capture system for administrative reporting of enterprise using Internet. Most often, the IDCS provide electronic questionnaires:

- To be filled in on line, with very poor possibilities for validation of the data.
- To be filled in off line, without possibilities for validation, linking data from computerised company accounts to the questions or other "intelligence" in the questionnaire that will be maintained automatically, to be sent by e-mail.

Both alternatives are not sufficient for an effective and efficient data capture system for administrative reporting of enterprises. This section of the paper especially focuses on the additional requirements for enterprise reporting.

In figure 1 the most relevant parts of the infrastructure of the data provider and the data collector are shown, as far as the data collection process is involved. The data collector has its (statistical processing) system to produce statistical figures based on the data collected from the data providers. The data collector also has its questionnaires, standard classifications and register of respondents.

The data provider has its own ICT environment, in which the automated accounting systems (most often the source for the data to be collected by the data providers), the operating systems used, as well as the standard (accounting and Office) software play an important role.

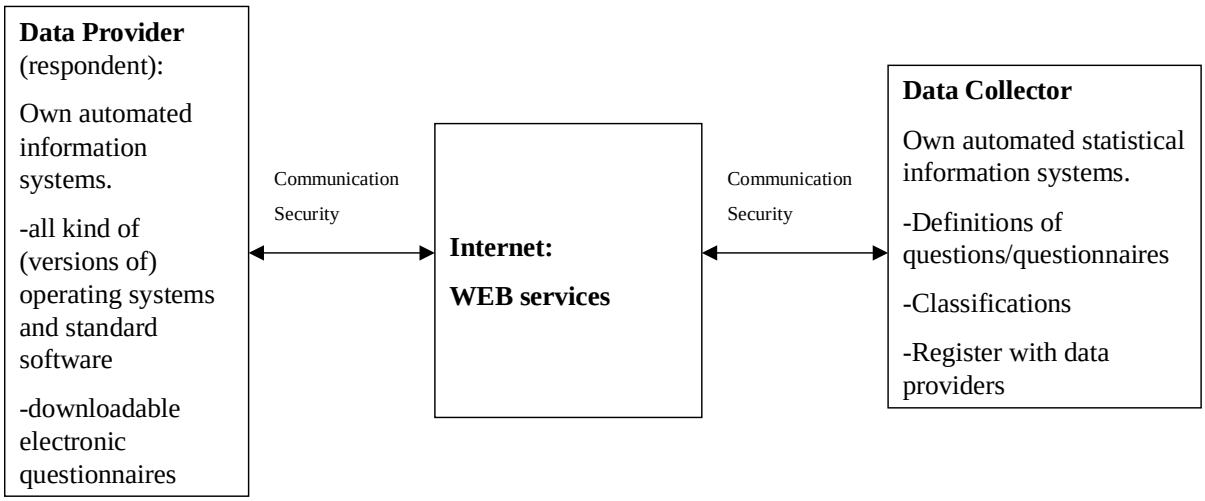


Figure 1. Internet and a data capture system

The Internet server should not be integrated within the technological infrastructure of the data collector. The mean reason is security. The technological requirements for the Data Provider (respondent) should be minimised and standard software should be used. The mean reason is here to cover as many respondents as possible with the data capture system.

The main requirements of the information system at the data collector are:

- A tool should be available to create the questionnaires and maintain the questionnaires.
- Possibilities to use the same questions in more than one questionnaire (harmonisation, standardisation, lowering the burden, etc.)
- Using standard technology for making and maintaining the questionnaire.
- A database in which data on the questions and questionnaire as well as data of the respondents (data providers and observation units) and the answers are stored.
- A function should be available to assign the questionnaire types to the respondents.
- The answers on the questions should be stored in the database, so that they are still encrypted but nevertheless connected to the metadata and the observation unit.
- Batch functions should be available to “upload” data from the Business Register to the database on the Internet server as well as to “download” the answers from the questionnaire to the statistical information system.

The main requirements of the software at the data provider are:

- The questionnaire should be downloadable from the Internet server.
- One should be able to fill in the questions:
- Manually (just typing the answers)
- Automatically by initially defining a link between the administrative data from the automated accounting systems and the questions.
- Partly manually and partly automatically.
- Changes in the administrative data or in the questionnaires should not influence the already defined links between the administrative data and the questions.
- Using standard software.

- The main requirements for the data transport over the Internet are:
- Availability of software to be sure that the intended parties are communicating with each other.
- Security of the data on the observation unit and the questionnaire.
- Security of the answers.
- The format of the data should be easy to transform to the format required by the (statistical) information systems of the data collectors.

3. The place of the data capture system in the Statistical System

An information system that meets the requirements for a data capture system using the Internet, combines the power of the Internet with the upcoming standard of XML[1] and the "de facto" standard of Excel, as far as linking bookkeeping data and questions are concerned.

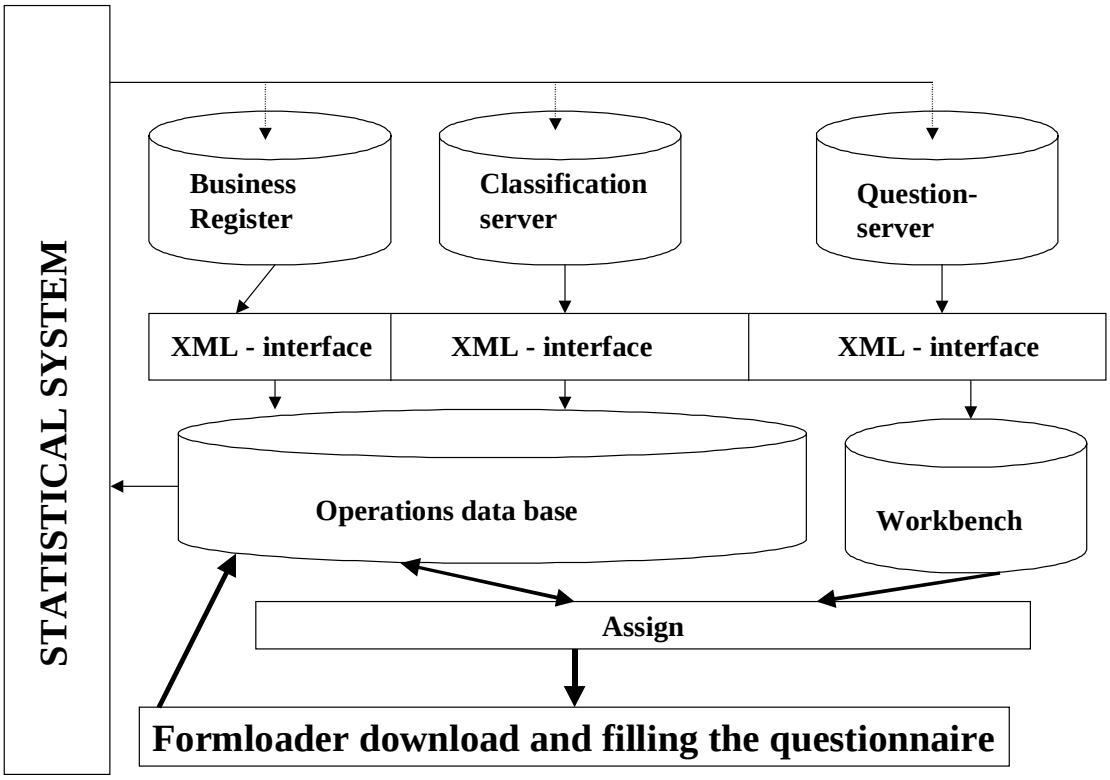


Figure 2. The data capture system and the Statistical System

There is one central part: a database with data and software to control the data collection process. The respondent (data providers) only needs an Internet connection and Excel software. The - by the "central part" - generated forms or questionnaires to be filled in are downloaded by the respondents via Internet and after transferring his own data to the form, the data are sent to the central part.

The data capture system should be easily connected to the Statistical System. This is possible by defining XML-interfaces between the data capture system and the related databases like: the businesses register, with data on observation units and data providers, the question database with meta data concerning the questions (name, definition, syntax rules, classifications used, etc.).

[1] XML is a de factor standard for communicating structured data over Internet. XML is supported by Microsoft inc.

The data capture system is based on international standards and can be connected to the production system as a black box. Because of this, the implementation can be fast and easily. Implementation can be done in an elapsed time of weeks to months, and not of years or tens of years like we are used with integrated standard package solutions.

The main benefits for the respondent (enterprise or VSE) of the use of the system are:

- Data can be easily transformed from the information systems of the respondent into the formats required by the data collector. The software enables one time installation of transformation rules and by that, avoiding the routine tasks of filling in the questionnaires.
- Fast learning by the respondents to transform their data into those required by the data collectors: most people at the administration know how to use spreadsheets like Excel.
- Internet, as a carrier of the messages between the parties involved, is used in the most effective way, within the required specifications of the system.

4. The structure of the data capture system

The data capture system itself consists of three parts: Providers software, Workbench and Operations. The data collector controls two parts: the Workbench and Operations. This software is installed on: a workstation of the Collector and on an Internet server. The Data Provider controls the third part: the Formloader, this software operates on the PC of the Provider. The picture below shows the connections between the Internet server, the Collector PC and the Provider PC, as well as the main communication flows (security excluded).

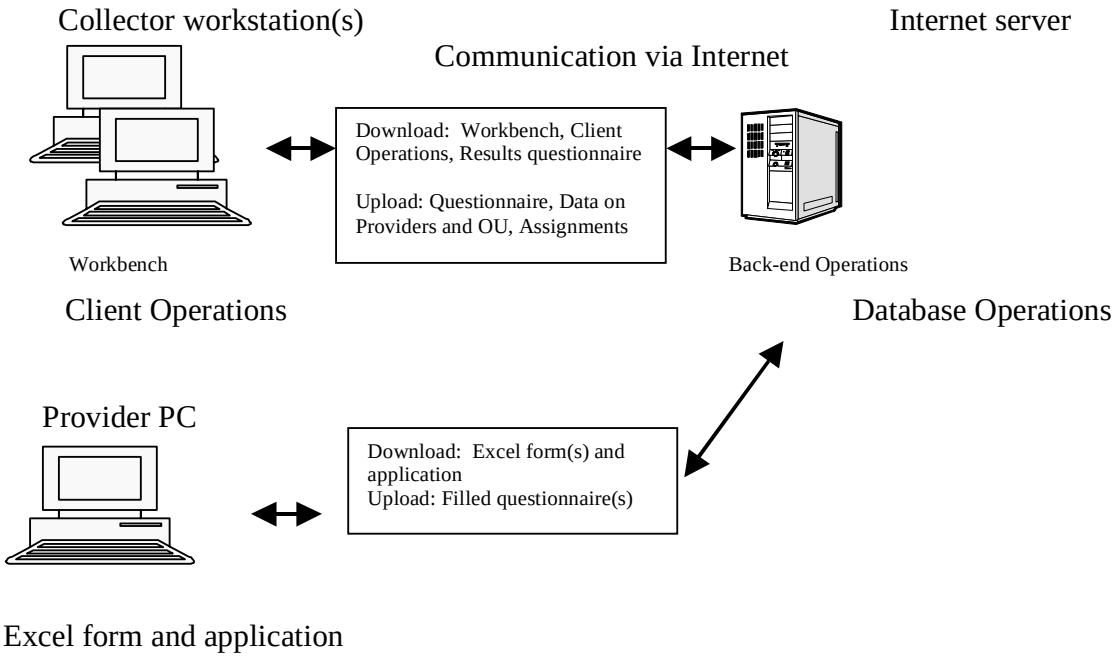


Figure 3. The structure of the data capture system

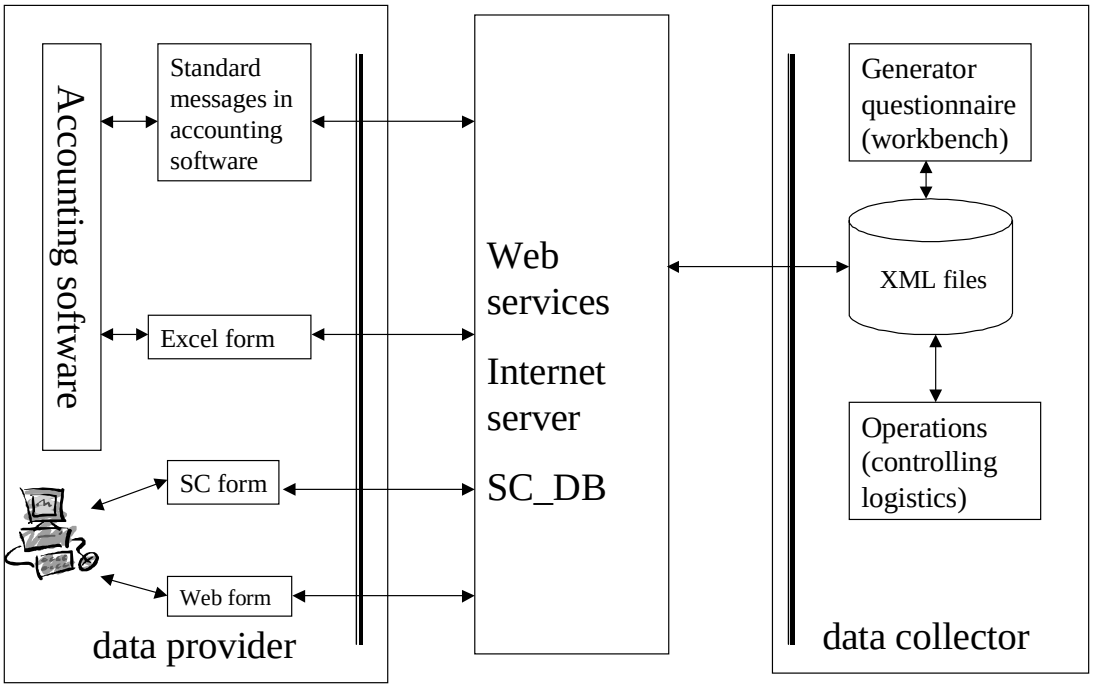


Figure 4. Data flows between data collector and data provider.

The questionnaire referring to a certain period of validity

This means that the questions in the questionnaire refer to a certain date (for example the end of a month) or to a period.

- The period of validity together with the type of questionnaire defines the questionnaire.
- One questionnaire comprises a number of defined questions, each question refers to one period or one reference date and each question should be answered according to defined syntax rules.
- An answer to a question can be asked for according to a certain code system (for example a standardised chart of accounts or an international product classification).

HTML Form

This form can be retrieved in a browser. It suits small and simple questions.

- SuperForm. This is a form in a proprietary format (written in Delphi). The main advantage is that the respondent does not need standard software (like Excel or Acrobat) to fill the form off line. With this form, complex questions can be asked and answered.
- Excel form. This is a form within Microsoft Excel. One of the advantages of this form is that no programmers are needed to link data from different systems. In each organisation, one can find employees with knowledge of Excel. They use Excel to edit and transform data, produced by the computers. They are now also able to use Excel to send the data. This is the solution for the organizations that have to collect data that are stored in different computer systems (like ERP and CRM systems).
- SCOOL library. SCOOL stands for SuperCollect Open Object Library. It is a set of specifications for publishers of standard (accounting) software to enable them to “link” data items from the databases of their standard software to the questionnaire types of the data collectors. This can be considered as an “invisible” electronic form, allowing direct linking to other software systems. The only action needed by the software systems is calling a few functions and the data are transported automatically to the server. This library enables a fast and efficient production of specific software for data capture. For example: software for the data flows between suppliers and shops, and between maintenance personnel and the planning department.

- After the initial installation of the software, the data provider downloads his overview of the questionnaires to be filled, using a password given by the data collector. The data provider selects the questionnaire and the completion of the questionnaire can start.

The Provider

The provider: in the form for the next period, the data from the previous period will be inserted in the Excel form. The provider only has to change data that have been changed.

- The collector: the lower cost and better quality of the receipt of data by the Internet.
- The provider can decide to make full use of the functionality of Microsoft Excel.

The system will check the consistency of the content of the form.

- After the instruction to send the data, the data will be encrypted, translated in XML and send to the data collector.

Security

- **Digital certificates:** The collector and provider identify themselves with a digital certificate. All data will be digitally signed. This process is automated, so neither the collector nor the provider of the data are aware of these preventive measures.
- **Encryption:** Different kinds of encryptions should be used, each with respect to a specific situation.
- Public-private key encryption. This is also called asymmetric encryption. This method is used:
- From Provider to Collector. The Collector generates both keys. The Collector sends the public key to the Provider. The Provider encrypts the responses with the public key, the Collector can decrypt them with his private key.
- From Internet Server to Provider. The Provider generates both keys. The Provider sends the public key to the Internet Server. The system on the Internet server encrypts the questionnaire and directives with this public key. The Provider can decrypt them with his private key.
- This method of encryption is very slow. So after authentication a symmetric session key is generated. This symmetric session key is used once for the data exchange between provider and collector.
- Symmetric encryption. This is used for the communication between Operations and the Internet Server. Both sides have the same key. This method is very fast, which is necessary because a lot of data can flow between Operations and the Internet Server.

5. Conclusions

The benefits of SuperCollect may be characterised as follows:

- Enterprise surveys: lessening their administrative burden (see results of the Teler project [1]).
- Household surveys: not only by applying CASI but also using the system in the CAPI interviews.
- Better response because of lower burden on (and emotional resistance by) the data providers in filling the forms.
- Better quality of the response by missing the failures of filling the paper form and subsequent data entry, including checks on syntax and control on content in the electronic questionnaires and the direct link with the accounting data in the information systems of the data providers.
- Considerably lower cost in the preparation of the statistical processing of the data. Less people have to be involved in the data entry and data correction procedures.

[1] In the TELER project the lessening of the administrative burden for the enterprises proved to be close to 80% when automated links between questions and accounting data could be provided.

- Less dependency of uncertain statistical valuation methods (imputation, automatic correction, etc.).

The data provider gets easier opportunities to translate its internal data into the required answers on the questions, which will lessen their administrative burden. Besides that, the data provider only has to worry about the meaning of the questions as soon as they are renewed. Only the first time he has to translate his own data source into the requested variable on the questionnaire. If the data provider uses standard software of software publishers that adapted SCOOL, the data provider even do not have to worry about the translation of his data into the questions at all.

References

Stol H R (1999), The use of models as a co-ordination mechanism, in *Research on Official Statistics*, EC 1999, ISSN 1023-098x

Stol H.R., An architecture for EDI in business surveys based on the use of Blaise III, in Essays on Blaise 1993, OPCS, London, 1993.

A full overview of the content of the SuperCollect system and the organisation of SuperCollect can be found on www.supercollect.com

Stol H.R., The Teler project and forthcoming developments in data capture, ETK ’99

Stol H.R. Metadata in data capture, Radenci conference 2002, Ljubljana, Slovenia.

About the Author

Hans Stol is General Director of SuperCollect b.v.

The author may be contacted at SuperCollect b.v. Noordeinde 20-22 2611 KH Delft, The Netherlands. Tel. +31 15 213 0808, or email stol@supercollect.com

.

SURVEY AND STATISTICAL COMPUTING IV.
Edited by R. Banks *et al*.

Building Standard-Based Survey Software

Mark Cameron & John Weisberg

Abstract

Techneos Systems is the developer of the Entryware system for data collection using mobile computers. The current release of Entryware products (v4.0) provides both case data and metadata support for the SPSS MR Data Model. This paper will examine two areas related to standards for survey software:

1. What issues must developers address in deciding whether to support survey software standards, and which ones to support?
2. What challenges have we encountered in our first iteration of support for the SPSS MR Data Model?

The first part of the paper will focus on business issues relating to adoption of standards, based on our experience in developing and marketing survey software. The second part will be more technical in nature, highlighting experience gained from implementation of both Metadata and Case Data Source Components (MDSC and CDSC, respectively) for the SPSS MR Data Model.

1. Introduction

One of the most common challenges we encounter today relates to interoperability. Nowhere is this challenge more evident than in human communication. Whether we are communicating via the Internet, cellular telephone, “land line” or face-to-face, many roadblocks stand in the way of seamless communication. These roadblocks largely stem from differences in technology, culture and language.

Adding to the “Interoperability Challenge” is that businesses, in the face of heavy competition, are under pressure to develop proprietary processes and technology to set them apart from their competitors. In these difficult economic times it becomes apparent that we must focus at least as much energy on working together with other companies—many of whom we would traditionally view as competitors—as we do working to differentiate ourselves.

In the market for survey software, the competitive race for creativity has generated a significant challenge for interoperability. Many software developers have excelled in specific areas of survey software, and seemingly countless firms have emerged as providers of “integrated survey solutions.” But while each vendor provides software that meets the needs of some researchers, no single application can meet the varied needs that have arisen within this industry. Just as we have each found different ways to solve many of the same problems, we must now find ways to share in more commonly accepted solutions.

It has long been evident that both buyers and providers of survey software would benefit from a more cooperative approach to delivery of software solutions. But with so many companies firmly entrenched in their own proprietary products and processes, the promise of interoperability faces many challenges. The essence of the "interoperability challenge" has been thoroughly covered by others over the past few years, so we will focus on those issues that Techneos Systems has encountered attempting to "plug in" to the existing marketplace for survey software.

Techneos has long been an advocate of standards for survey data interchange, though like many other firms we have had limited resources to allocate to the development and/or implementation of standards. We recently implemented support for the SPSS MR Data Model, which we believe will become the backbone for data interchange within the market research industry. While our decision to implement this model had strategic benefits due to our distribution alliance with SPSS MR, our decision to implement it was primarily based on the tactical benefits that it provides.

2. A Standard Amongst Standards?

The "multi-mode" dream has existed for many years in the survey research arena. Most researchers believe, and rightfully so, that it should be possible to design questionnaires using their choice of authoring tools, then deploy questionnaires to any data collection methods, or "modes", and to effortlessly combine data from these different modes using a common system for analysis and reporting.

Since our inception in 1995, Techneos Systems has focused on providing mobile data collection software. In the context of industry terminology, we specialize in MCAI—Mobile Computer Assisted Interviewing (or more specifically, MCAPI and MCASI for Personal and Self-completed surveys, respectively). As a niche player in an already established industry, we have always felt the need to focus on being the best at MCAI, and to work with other companies to provide multi-mode solutions.

We believe the optimal balance between competition and cooperation exists in the emergence of standards for survey data interchange. We have applauded the efforts of those companies and individuals behind initiatives such as Triple-S and AskML, and we have seen examples of the success which cooperation can bring to all parties involved. At the same time, we have found it challenging to decide which standards to support in a world of multiple—and to some degree competing—standards.

For a standard to be effective for us, it must possess the following qualities:

Rich functionality—as a small company looking to "plug in" to existing initiatives, we need to know that we will not outgrow the specifications of any standard that we adopt. While a "subset" approach to standardization works well in some circumstances, our vision calls for a "superset" approach that will allow us to provide best-of-breed components that work with a robust suite of products from other vendors.

Pervasiveness—in serving a wide variety of clients worldwide, we look to adopt standards which span geographic borders, and which reach a broad range of users and complementary technology.

Functionality

There are three areas of functionality that Techneos requires from a standard (or set of complementary standards):

Case data access—for data from our application to be used by other analysis and reporting tools in a multi-mode setting, there must be a common interface for accessing case data. It is essential that we are able to publish data for other tools to use, but since we are focused on data collection our software does not require access to data from other applications.

Metadata read/write access—with so many different authoring tools in use, and no single tool emerging as a standard for multi-mode data collection, it is important that our authoring tool can share metadata with other authoring tools. Even more importantly, we believe that a standard metadata interface is essential to define the context for case data.

Scripting/routing logic—Techneos, like most other survey software firms, has developed our own proprietary scripting language for regulating interview flow. For multi-mode data collection to become a reality across vendors, a common scripting interface is required.

For a standard to meet our needs it must also be equipped for internationalization—i.e. compliance with single-byte and multi-byte character sets—and it must be flexible enough to allow for customized properties that apply to only a single mode of data collection.

Pervasiveness

Even the richest specification will not provide value unless people use it. Like the infamous "beta vs. VHS" battle, which saw an inferior video standard become pervasive while a superior standard quickly became extinct, any standard only really exists when it is widely accepted.

To be truly pervasive a standard must cross many boundaries. It must be accepted in most, if not all, regions of the world; it must be recognized by large and small organizations; it must be adopted by a broad range of users and it must be implemented by all leading software providers. No standard will become pervasive overnight, but to justify its adoption a standard must demonstrate a likelihood of pervasiveness across all of these boundaries.

Who's backing the standard?

Many attempts have been made to develop software standards over the years, and relatively few have succeeded. To cite a relevant example, countless efforts have been made to develop cross-platform development tools that allow "build once, deploy anywhere" functionality. A few products, such as Java and Kylix, have seen limited success as cross-platform development tools. But by and large, competition between platform providers such as Microsoft, Apple and Sun (or in the mobile world, Microsoft and Palm) stand in the way of real progress toward this goal. Although Java was introduced as an "open standard", both its successes and its failures can largely be attributed to the fact that it is commercially backed by Sun Microsystems. It can be said that there are "open" aspects of Java, but few would argue that it is truly an open standard. Similarly, the emergence of .NET is a clear indication of Microsoft's goal to develop and control a standard platform that competes directly with Java.

Neither Java nor .NET will ever provide the type of open source, cross-platform standard that many people dream of. What they do provide, however, are practical backbones upon which businesses—both users and developers of software—can develop productive software tools to meet their needs. The success of each platform is tied directly to the backing that it has been provided by a world-class organization. When we look at standards for survey data interchange, we must also look at who is backing it and what resources are being allocated to its development.

Priorities for implementation…

In a perfect world, Techneos would see our software achieve interoperability with as many other products as possible. Most importantly, we are committed to providing a "best of breed" mobile data collection tool for use in a multi-mode environment. While proprietary interfaces are an option to bridge the gap between our software and other applications, we prefer to focus our limited resources on implementation of one or more standards that will meet our needs.

Techneos has monitored the progress of both the Triple-S and AskML standards. Both of these initiatives have been covered in detail by other authors (who have significantly more insight into these initiatives than us). Thus we will not delve into their specifics, but in the context of this discussion we will say that we view Triple-S and AskML as important initiatives in the survey software arena. Techneos has great interest in both standards, and we believe that adherence to both standards may become an essential part of our product roadmap. However, we have chosen to focus our immediate efforts on implementation of the SPSS MR Data Model, for the following reasons:

The SPSS MR Data Model promises to meet our needs for both functionality and pervasiveness. The Data Model already provides a robust framework for sharing case data, and a sufficient metadata specification to provide analysis context for case data. The model does not yet support scripting of interview logic, but SPSS is committed to extending its robust mrScript language for this purpose.

With the strong commercial backing of SPSS, the MR Data Model also shows promise of becoming a pervasive model for both researchers and software providers alike. Although "organic" standards like Triple-S and AskML offer a more open approach than the MR Data Model, the development of these models is limited by the mostly volunteer resources that are allocated to them. In contrast, SPSS has allocated a huge amount of resources to development of the SPSS MR Data Model, which is rapidly gaining support amongst research companies.

What functionality is most important?

We have received significant feedback from our customers about standards for data interchange. Though informal in nature, the feedback has consistently indicated that case data integration is the highest priority for implementation. Intuition may suggest that metadata integration is the highest priority, and our initial view was that a common authoring tool was the key to multi-mode research. However, we have learned to differentiate between metadata that is required to provide analysis context and metadata that is required for seamless multi-modal deployment. While we are ultimately committed to working within a multi-mode environment, we understand that the most important use of metadata is to define the case data that is collected for a project.

The justification for putting case data first is that only a small portion of the total time and energy required for most projects is allocated to "programming" a questionnaire. By contrast, reporting and analysis of case data represents a significant portion of time for the most projects. This assertion is made notwithstanding questionnaire design, which is generally done prior to the "programming" process. As a producer of data collection software, it is our goal to minimize the impact that a specific mode of data collection—in our case the limited form factor of mobile devices—has on questionnaire design.

The SPSS MR Data Model meets our current needs for case data interchange, and it is well on the way to providing the metadata and scripting support that will be required for seamless multi-modal interviewing. While we grow *with* those parts of the model that are still under development, we are also challenged to grow *into* the extensive functionality that SPSS MR has already implemented.

3. Implementing Support for the SPSS MR Data Model

Once we made the decision to support the SPSS MR Data Model, we formulated the following high-level implementation strategy:

1. Determine a mapping scheme between our proprietary metadata model and SPSS's Metadata Document (MDD) format;
2. Implement a Metadata Source Component (MDSC) to convert our metadata into an MDD object;
3. Implement a Case Data Source Component (CDSC) to make our case data accessible to other Data Model compliant applications;
4. Implement a component to map MDD data back to our proprietary metadata model;
5. Add user interface to our authoring tool, Entryware Designer, for importing and exporting metadata via the Data Model; and
6. Integrate MDSC and CDSC components into our software installation process.

Mapping Metadata

The MDD format is a robust and extensible object model for representing virtually any type of metadata. Though I'm told that it has limitations in areas such as hierarchical data, we did not encounter any shortcomings in the MDD spec that couldn't be overcome by use of custom properties.

For our first iteration of metadata "export", we chose to minimize our use of custom properties. Since few applications would likely make use of Entryware-specific properties, we focused on exposing metadata that provides context to case data. This primarily consists of questions, variables, and information about categorical data (i.e. native values and labels). We also used custom class objects to map environment information such as respondent numbers, timestamps and device identifiers.

We were pleased to find that a substantial portion of the MDD specification could easily be mapped back to our proprietary metadata model, for both "export" and "import" purposes. We were able to directly map simple question types such as messages, single-response categorical questions, and open-ended (numeric or text) responses to our native data types. Most categorical loops and grids were also easily converted to our data model, though we had to bypass some more complex categorical loops and numeric loops, which Entryware does not support. Questions in an MDD document that cannot be supported by Entryware are filtered out during the "import" process.

We did not attempt to implement routing logic using the MDD specification. Though limited routing logic could be stored within the MDD and complex routing could be implemented using custom properties, we have chosen to wait until SPSS adds mrScript integration to the metadata document.

We also chose not to utilize the Data Model's versioning control, though we expect to use it in the future. Although the Data Model provides extensive capabilities with regard to version control, a great deal of work would be required to map our proprietary version control to that of the Data Model.

Our greatest challenge in mapping metadata was to understand the object hierarchy of the MDD specification. When we embarked on Data Model implementation, the available documentation was sparse and the sample code overly complex. SPSS has significantly improved and expanded this information, but it is still quite a task to fully understand the metadata hierarchy. In many cases there are multiple ways to achieve the same goal, and the documentation provides only limited help in deciding which options to choose. If we can offer one piece of advice to anyone implementing MR Data Model support, it is to make use of the discussion forum that SPSS MR provides. Anytime we have asked for help, we have received a prompt and thorough reply.

Implementing the MDSC (Exporting Metadata)

Once we mapped our metadata format to the MDD model, the exporting process was relatively straightforward. We did however encounter a few notable issues along the way:

Native values—the Data Model stores each category in a questionnaire with a unique value, whereas Entryware uses native values that are only unique per category list (i.e. per question). The Data Model documentation for handling native values is somewhat convoluted, but now that we understand the process we feel that it is a flexible approach to handling categorical data.

Multiple-response categorical data—a bit of extra work was required to deal with multiple-response categorical questions. Entryware represents each response item as a sub-question, mapping each data value to a single variable with either a "Yes" or "No" response value (for selected or deselected, respectively). The Data Model uses set notation, where each item's value is to be returned within a set.

> *E.g. in Entryware, selecting "Blue", "Green" and "Red" from a list of colors would result in 3 variables: BlueVar = "Yes", GreenVar = "Yes", and RedVar = "Yes", with all unselected items being saved in separate variables with "No" values. Using the Data Models' set notation, this data is represented as ColorQuestion = {"Blue", "Green", "Red"}.*

As with the native values, we have found set notation to be a useful and flexible way to deal with categorical data.

Grids vs. Arrays—our first build of the MDSC used arrays to represent loop data, i.e. sub-questions that each require either a categorical, numeric or open-ended response. Although arrays can be used for this purpose, we found that grids provided a much more direct mapping to our model. We were not able to fully test this functionality until we had implemented both read and write support for the MDD format, as this allowed us to export questionnaires and re-import them to Entryware, then compare the original questionnaire to the imported one.

Once we had developed our MDSC to convert our metadata into an MDD document, the Data Model provided a simple function for saving the MDD document (in memory) to a data file.

Implementing the CDSC

Conceptually, case data support is much simpler than metadata support. Much of the work required to implement a CDSC is already in place within the MDSC. However, in practice we found a few challenges exposing Entryware case data to the Data Model:

Performance—our first iteration CDSC is effective, but admittedly quite slow. There are a number of causes for slow performance, but we now see many ways to improve the speed of our CDSC in subsequent releases. The two main contributing factors are:

1. Entryware was designer to process data in batches, not in a cumulative fashion. We had to develop a new data processor to support the CDSC, and many optimizations have yet to be implemented for this new component;
2. The MR Data Model has to support a broad range of case data formats, and it has to apply metadata context to each piece of data it receives. As a result its method for requesting data during queries is divided into very atomic calls. In simple terms, this means that each cell—i.e. each variable for each respondent—requires significant processing overhead.

Table Spanning—when we developed the data processor for Entryware Designer, it was built to export data from our proprietary binary format directly into either ASCII data or a formatted SPSS dataset (i.e. ".sav" file). Because both of these formats support record structures with thousands of variables, no additional logic was required to manage case data export for large questionnaires. Our CDSC

component uses a relational database to manage processing and exporting of case data. Since most relational databases support only 256 to 1024 variable/fields per record, additional logic had to be added to manage "table spanning" for larger questionnaires. This logic has also impeded performance of the CDSC, but we have now established ways to overcome the effect of table spanning on processing speed.

It is important to note that the MR Data Model does not simply support *data export* to a common format. It provides an *interface* for applications to expose their case data to other Data Model compliant applications in a consistent fashion. While this approach is more complicated for developers, it is far more powerful than a simple export specification. By implementing both a CDSC and a corresponding MDSC, we are exposing our case data so that researchers can seamlessly integrate it with data from other sources. And when combined with mrScript and mrDataManager technologies, CDSCs can be used to automate analysis and reporting in a multi-mode data collection setting—regardless of what authoring tools were used to create the questionnaires for each mode.

Importing Metadata Documents (MDD)

Importing of MDD documents is essentially a mirror of the process for exporting via the MDSC. As mentioned above, much of our testing and fine-tuning could only be done once we had implemented both import and export functionality for MDD files.

Putting the Pieces Together—Adding the User Interface

For user interface, we only had to add "Save As Other" and "Open Other" to the file menu in Entryware Designer. The Data Model does all of the work required to transfer data to and from other compliant applications via their MDSCs. Because each compliant application is required to register its Data Source Components (both MDSCs and CDSCs) with the Data Model, very little additional work is required to open from and save to other compliant applications.

In our latest version of Entryware Designer, v4.0, we implemented saving as MDD files, and opening from MDD files and any other file types registered with the Data Model. We did not implement "save as" functionality for other file type, but that will be added to a future version of Entryware Designer.

Installing the Components

One of the most important—and most painful—aspects of any software development project is to ensure that users are able to seamlessly install the product. We managed to get our DSCs to install properly after some trial and error, a bit of "reverse engineering" of the Windows Registry (based on other Data Model compliant applications that were installed), and some timely help from SPSS MR. Nevertheless, this important part of the software process has not received its fair share of SPSS's documentation of the Data Model.

4. Conclusion

It is becoming increasingly important for software developers to work together, yet in many ways we continue to widen the gaps between ourselves as we each work to establish our own competitive edge. With so many survey packages on the market, each using its own proprietary data structures and processes, true interoperability cannot be attained through one-to-one integrations between developers. It is essential that developers and researchers alike rally behind standards for survey data interchange.

Standards offer perhaps the best example of the old adage that "perception is reality," for a true standard exists only if it is widely accepted. The well-established Triple-S standard has demonstrated real value for both software developers and researchers, and emerging standards such as AskML offer promise for a future of open communication and interoperability.

While there are numerous initiatives to develop standards for survey data interchange, Techneos believes the standard that is best suited to our needs, both currently and those that we foresee for the future, is the SPSS MR Data Model. The MR Data Model provides a robust framework for sharing metadata and case data, and in conjunction with mrScript and mrDataManager technologies, we expect it to provide all the functionality that we—and our customers—will require for true multi-modal data collection.

Techneos has completed our first iteration of MR Data Model support. Though we have encountered some minor issues during implementation, we have found the data model to be robust and flexible enough to handle every challenge we have encountered.

I am pleased to report that we have taken our first step toward interoperability with other survey software applications. As standards continue to become more functional and more pervasive, software developers who do not support open interchange of survey data will face increasing pressure from a growing network of inter-connected applications.

I look forward to seeing the multi-mode dream finally become a reality…

About the Authors

Mark Cameron, BSc, and John Weisberg, BA, are co-founders of Techneos Systems Inc, located in Vancouver, Canada. Pioneers in the field of Mobile Computer Assisted Interviewing using handheld computers, Mark and John have been developing and using handheld survey software since 1991. Their software has been used to conduct over a million case records, on devices ranging from DOS-based handhelds to the latest Palm OS products. Both Mark and John have been closely monitoring the development of standards for survey data interchange since 1995.

SURVEY AND STATISTICAL COMPUTING IV.
Edited by R. Banks *et al*.

First Installation of a Virtual CATI Phone Room with Full Telephony Integration

Louis Tanguay

Abstract

A natural extension to a Web survey engine is a Web-based CATI (Computer Assisted Telephone Interviewing) system. Voxco has been pioneering this idea for many years, and several software providers now claim to offer such a solution. Although Voxco's Web-based CATI offering was in theory ready as early as 1999, we only released it as a commercial product recently because we (and potential clients) felt that without appropriate telephony integration, a virtual CATI call centre could not provide sufficient control over field operations.

Since Voxco is both the author and owner of our own telephony solution, it was totally logical for us to create a Web-based CATI system featuring integrated telephony and predictive dialling.

A possible organizational intermediate step in implementing a geographically unfettered Virtual Call Centre (VCC) with telephony integration is a Web-based CATI with telephony gathered in a single location. Voxco's telephony solution PRONTO already offers the remote interviewers function, so the extension to the real VCC with home interviewers is simply one of logistics for the client; all the software components being already in place and validated.

This first step has been successfully implemented with the installation in September 2002 of a 65 station browser-based CATI phone room with predictive dialling at one of our clients in Montreal. This paper describes the original implementation, the technical issues that have arisen, and how they were solved.

Keywords

CATI system, telephone surveys, VCC.

1. Introduction

In recent years, there have been many paths leading software providers to the development of a Web-based CATI system. The two most common ones adopted in the industry depend mostly on the age of the company. Younger companies typically first develop a Web-based survey engine, and then realize that it could very well fill the main functions of a CATI system – the recent debacle of the Web economy certainly steered some of them in that direction. Older companies are more likely to have developed both a CATI system and a Web survey engine in parallel – hopefully with a common engine kernel, in order to support existing installations and clients. This is the case for Voxco, nevertheless, since the very beginning of our Web survey engine development (1998), it was clear to

us that it should also be extendable to a Web-based CATI system. Several software providers now claim to offer such a solution. Although Voxco's Web-based CATI offering was in theory ready as early as 1999, we only released it as a commercial product recently because we (and potential clients) felt that without appropriate telephony integration, a virtual CATI call centre could not guarantee a sufficient level of control over field operations – due to the fact that interviewers may work at a distance and may not be grouped in a single location.

The term *telephony integration*, in this context, refers to a set of functions which guarantees a sufficient level of control over the dialling and conversation processes, as well as a consistent relationship between this process and the corresponding survey data. In order to achieve this control, we believe that the minimal set of functions should be:

- automatic dialling of the actual phone numbers in the sample file;
- accurate and reliable timing of the actual interviews;
- audio monitoring of any conversation.

A more complete set of functions – let us call this *full telephony integration* - that one would ideally expect from telephony integration includes:

- automatic classification of non connected calls (*call progress analysis*) – with full logging of this activity;
- predictive dialling (i.e. intelligent optimization of the telephone resources – and the associated productivity gains);
- conferencing and call transfer functions.

Telephony integration, by virtue of its audio-monitoring capability, offers the required level of control. Even when an interview is not monitored, one can be sure of which phone numbers are actually dialled, as well as the results of all non-connected calls. Since Voxco is both the author and owner of our own telephony solution, it was totally logical for us to issue a Web-based CATI system featuring integrated telephony and predictive dialling.

Several other significant benefits arise from full telephony integration:

- Centralized billing of all calls: since all calls are issued from the telephony server, they are billed to the central location; in most of North America, local calls are free, and therefore, at least in large urban areas, the interviewer connection to the telephony system generates no extra cost.
- Centralized audio recording and play-back capabilities; to play jingles, or record the verbatims digitally, a centralized telephony server offers an efficient and economic solution. Otherwise, one would have to manage different soundboards, soundfile downloads or uploads etc.
- Call transfers to supervisors or skilled interviewers after a screener. Although this function is highly specific, it can be economically interesting, especially in fields where costly skilled interviewers are involved (pharmaceutical, medical studies).

2. Beta Site

After intensive alpha-testing and simulations, a first small-scale beta-site of our Web-based CATI system with telephony integration was implemented in the Summer of 2002. They were five to ten browser-equipped workstations at a client's site. The tests were controlled ones, with Voxco staff permanently on site, and a single CATI project running at a time. All changes to projects were made by Voxco, to avoid any external interference. The beta-test lasted for more than a month and, in spite of the normal problems to be encountered in such a new environment, and after several cycles of

correction and re-testing, proved to be successful. This convinced us that our product was ready for delivery on a larger scale, and in a freer environment.

3. Installation of a First Site

Regardless of purely technological issues, the logistical and organizational changes implied by a CATI VCC made it hard, however, to find clients ready to make such a move. A possible organizational intermediate step in implementing a geographically unfettered VCC with full telephony integration is the installation of Web-based CATI with telephony gathered in a single location. Voxco's telephony solution Pronto already offered the remote interviewers function, so the extra-step to the real VCC with home interviewers is simply one of logistics for the client; all the software components being already in place and validated.

This first step was successfully reached with the installation in September 2002 of a 65 station browser-based CATI phone room with predictive dialling at one of our clients in Montreal. The company, CROP, is a full service research house, offering social, political, and marketing research. It is one of the oldest research companies in Canada.

Although in the initial installation all interviewers work on site, the client's plan is to gradually increase the capacity of their phone room. This could be done by adding interviewers working from home. Due to the flexibility of implementation allowing separate telephony servers and survey servers in any combination, possible integration of other sites with the company's partners across Canada is also being considered.

4. Technical Description

Without going too deeply into technical details, a short overview of the system components will help the reader understand the overall context of this implementation. The three main components (Web server, Telephony server and Interviewer workstations) can each be in separate locations, and repeated if needed in different combinations (e.g. one Web server, with two Telephony servers and interviewer workstations in two separate locations).

- A *Web server* (Interviewer VCC) runs the CATI questionnaires. This is a DLL extension of IIS, under Windows 2000. It dynamically generates HTML pages, and saves the response data. It performs most of the sample management functions including quota incrementation etc. It also generates those HTML pages which are specific to a CATI workstation, such as a list of personal call-backs, appointment calendars etc.
- On the same PC as the Web server also runs a service called the *Call Server*. This service draws on the sample database and dispatches the cases to be called to the Telephony server, according to the quota rules, call-back rules, time zone, case priorities etc. It also records the call results of all unconnected calls, and schedules the next call to be done if necessary.
- A *Telephony server* (Pronto) performs the actual dialling. This is a service on a Windows 2000 server, with the appropriate specialized hardware (telephony cards). It manages external and internal phone lines. It runs prediction algorithms based on response rates and call durations for each CATI project. It also performs the call progress analysis and signal detection. Pronto communicates with all other components of the system through a CORBA interface. Since CORBA is a protocol which runs over TCP-IP, all messages can be routed through the Internet.
- Each *interviewer workstation* is equipped with a browser (Internet Explorer). The station communicates with the Web server via HTTP protocol (regular HTML pages). We will see below

that some JavaScript functions are also used to improve survey ergonomics. An ActiveX extension (*Pronto toolbar*) provides the interface with the Telephony server. This ActiveX extension communicates with the Interviewer VCC server through the current HTML page, and uses the CORBA protocol to invoke functions from the Telephony server. This provides access to all available telephony functions (login to the dialer, logout, receive a connected call from the dialer, dial a phone number, hang-up, play a soundfile, transfer a call, etc). Another ActiveX extension implements the video monitoring functions. For the telephone connections, interviewer workstations can be connected directly to the Telephony server or through a PBX, or remote interviewers can dial-in through a dedicated phone number.

5. Implementation Issues

Our client and ourselves have been faced with two types of challenges with this new installation: technical ones and organizational ones. The client's staff was well accustomed to the DOS version of our CATI Interviewer (more than ten years in use), and it was clear from the start that the new system should not sacrifice any functionality of the old one. Monitoring was one of these crucial functions.

Technical challenges

- *Monitoring*: the audio part of monitoring was already available with our telephony solution. However, the prototype of a video monitoring module over the Web, which we had already built using JAVA, was incomplete and still needed some improvements. Because of the relatively short implementation calendar, we decided to start by using the video monitoring components of our CATI system, in its LAN (Local Area Network) version, and to incorporate it as an ActiveX extension into the browser. This offered the advantage that the supervisor monitoring tool was already written, and supported the coupled audio and video signals. The drawback was that it would only work in a LAN environment, not yet supporting video monitoring over the Internet for real home interviewers.
- *Personal call-backs*: in addition to regular general call-backs, the Interviewer VCC system already provided support for personal call-backs managed by the interviewers. We did not expect, however, such an intensive use of this feature with this specific research company. Our original design was not intended for centralized support of hundreds of personal call-backs per interviewer. The situation was quickly resolved after the appropriate changes to that specific part of the code. This example simply demonstrates that real life always bring new situations that intensive in-house testing may never detect.

Organizational challenges

Testing and beta-testing do not provide complete insight on organizational issues, especially if the beta-testing phase involves a different organization, with different habits. The company was used to a different paradigm (i.e. the DOS version of Interviewer CATI). There are obvious differences provided by the GUI/ HTML-based interface, but more subtle ones due to the fact that users were switching from a decentralized application to a more constrained and centralized one. CATI users familiar with a main-frame type of application do not realize that there are drawbacks to centralization, because they are used to it, and do not even imagine that it could work otherwise. Of course, a more centralized approach also brings many advantages with the constraints that it imposes. Whether these constraints are perceived as positive or negative highly depends on the work environment and company size and structure.

- *Mouse-driven survey ergonomics*: initial beta-tests underestimated the reactions of interviewers to a mouse-driven interface, with standard HTML controls (radio buttons, check boxes etc). Real life showed that they felt slowed down, and this was reflected in productivity figures. They were used to keyboard shortcuts (entering the codes or using the arrow keys) for response selection, as

well as navigation through the questionnaire (escape/page-up/page-down etc). General acceptance of a system throughout an organization is a key to its success, and we reacted promptly to provide an interface equivalent to what the interviewers were used to, by the means of JavaScript embedded in the HTML page. Universality of browser support is a less important factor in a VCC than in Web self-administered interviewing, hence the idiosyncrasies of specific implementations of the JavaScript language becomes a minor issue. The interviewers responded with enthusiasm to these changes and their overall productivity returned to previous levels, all other things being equal.

- *More centralized environment*: switching from a station-centric paradigm to a server-centric one requires a bit of adaptation especially for supervisors, who have several different tasks to accomplish, and figures to monitor. The main challenge for them was to understand the new paradigm. Once this was integrated, new working procedures were developed and easily accepted.

6. Working From Home

Now that all the components of the system have gained full robustness and the confidence of the users, the extra step to real home interviewing is just one of logistics. All that is required for a workstation is a PC, a fast Internet link (even though a very small amount of information is transferred between pages, a one second delay can become costly when multiplied by the number of transactions involved in a day's work). Since an ActiveX is being used, the current version requires a Windows environment, with an IE browser. In addition to this, the interviewer requires a telephone line for the voice communication (it can be combined or separate from the Internet line). Interviewers can work in different time zones, the VCC server being aware of the relative time lag between the three parties: interviewer, respondent and the server itself. This can be critical when setting telephone appointments. The following image shows the interviewer main screen in VCC mode:

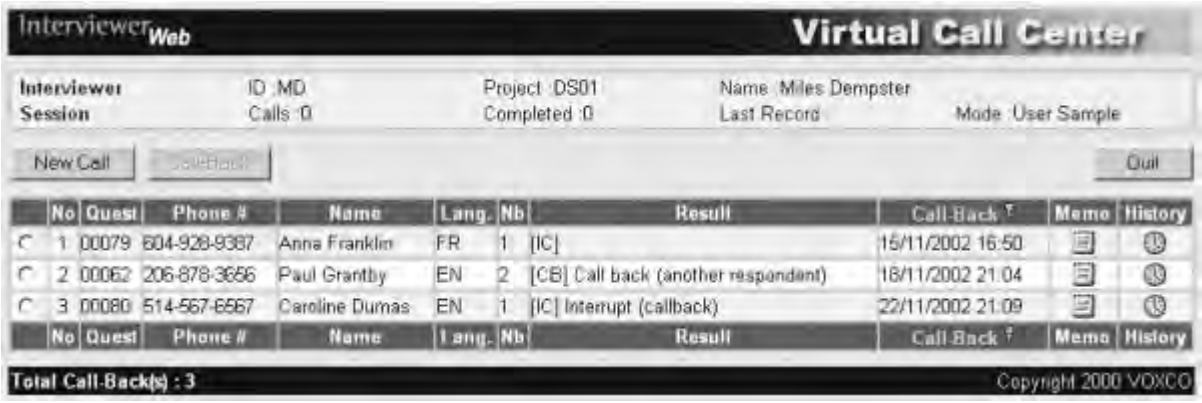


Figure 1. Interviewer Main Screen operating in Virtual CATI mode over the Internet

A typical sequence of events for a home-initiated interview is:

- the interviewer picks a URL that points to the survey (or group of surveys) on the Interviewer VCC server;
- he/she selects the survey (if needed) and logs in to the survey;
- he/she dials the dedicated phone number to access the Telephony server;
- after the Telephony server audio prompt, he/she uses the phone keypad to enter an identification key which matches the data station to the phone line;
- on the main survey screen, he/she clicks on a button that informs the Telephony server that the interviewer is ready to receive calls (assuming that we are in predictive mode – other modes are available) and waits for a live call to be connected and an interview to start;

- as soon as a call is answered, and it is this interviewer's turn to be served, the Telephony server transfers the call on to the telephone line; simultaneously, a message is sent to the browser, via the Pronto bar, instructing it to request the first interview page from the Interviewer VCC server together with the corresponding case;
- for soundfiles, when required, the interviewer uses the Pronto toolbar to access the record and playback tools: standard start and stop buttons. These tools can be activated automatically or manually, according to the questionnaire script.

7. What's Next

In the summer of 2003, Voxco completed the development of video monitoring across the Internet. Improvements are also being considered for the dial-in procedure, whereby the Telephony server could dial the interviewer's phone number, immediately after the login procedure. This simplifies the billing process, and increases security.

Use of voice over IP is a natural extension of the system, with a slightly more distant horizon. Bandwidth and sound quality are more of a concern in this case than are the technical implications of implementing this function, which only concerns the Pronto Telephony server.

8. Conclusions—The Benefits of a Virtual CATI Phone Room

A web based CATI system with telephony integration introduces major changes in many aspect of telephone data collection. By enabling working from home, without compromising quality control, it offers tremendous flexibility over many aspects of CATI operational workforce management. Just to mention a few:

- **Scalability**: increasing or decreasing the size of the virtual phone room depending on the workload is just a matter of contacting the interviewers. Deployment can be achieved in a matter of minutes.
- **Languages**: finding interviewers to work in foreign languages becomes easier, because one can recruit them directly in the target country, or subcontract CATI data collection while keeping full control on the sample management part of it.
- **Personal security**: due to extended shifts imposed by time zones, interviewer security is a problem in many North American cities. Working from home provides a solution to this problem.
- **Flexibility of working hours**: working from home extends the calling hours available. Moreover, predictive dialling makes it economically possible to place calls during low incidence hours.

About the Author

Louis Tanguay has a background in mathematics. Since 1978, he has been closely involved in all software product design and developments at Voxco, including three generations of a CATI system, a predictive dialler for the market research industry, and StatXP, a statistical tabulation package. He also has 25 years of experience as a consultant with Voxco's clients in survey data analysis using all multivariate techniques.

He can be contacted at Voxco, 1134 rue Sainte-Catherine Ouest, suite 600, Montréal, H3B 1H4, Canada. Tel: +1 514-861-9255, e-mail: louis.tanguay@voxco.com

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

We Seek Them Here, We Seek Them There. How Technical Innovation in Mixed Mode Survey Software is Responding to the Challenge of Finding Elusive Respondents

Tim Macer

Abstract

Increasingly, market and social researchers are turning to mixed mode surveys as a means to overcome a variety of research issues and constraints, not least the increased difficulty in achieving successful contact with respondents.

At a technical level, the central challenge of true mixed mode research is to combine both administered and self-completion interviewing across a range of devices and methods, each with its own inherent limitations, in ways that will minimise operational complexity and also the variability or 'modal effect' that is experienced between the different modes. Many commercially developed survey packages claim to offer support for mixed mode research, but in some cases are doing so with little consideration for the needs and requirements of mixed mode research.

This paper sets out the principal operational and technical issues that multimodal data collection software needs to address, examines specialist survey research software providers are responding to these challenges, and identifies innovation in this area. It concludes with recommendations on a minimum set of facilities or features that any mixed mode interviewing system should contain.

Keywords

Multi-mode surveys; mixed mode surveys; survey instruments; modal influence; survey software

1. Introduction

Survey instruments exist to take measurements. Not unlike the more physical instruments used by chemists or engineers, there is a need to calibrate survey instruments to ensure they yield consistent and accurate results. While it may be possible to test the instruments, destined for the chemist's laboratory or the engineer's workshop, in a controlled setting and confirm accuracy in absolute and predictable terms, the market and social researcher's instruments can only be tested in relative terms. The verification of survey instruments has been achieved through experience. But as researchers reach for mixed mode surveys to boost flagging response, the effect of combining modes into one survey on the results is relatively little researched and little understood (Cobanoglu et al 2001).

For many decades, the choice of modes for survey administration was restricted to face-to-face or telephone. Technological advancement today, allows researchers to choose from an ever-widening palette of survey modes, as figure 1 illustrates, many of which have evolved from each other

incrementally, as technology has improved. In this way, is possible to trace back almost all computer-assisted methods back to CATI, in the early 1980. It was not until 1996 that email and Web browser-based Internet surveys started to appear on the scene, but these have encouraged an explosion in the development of new modes.

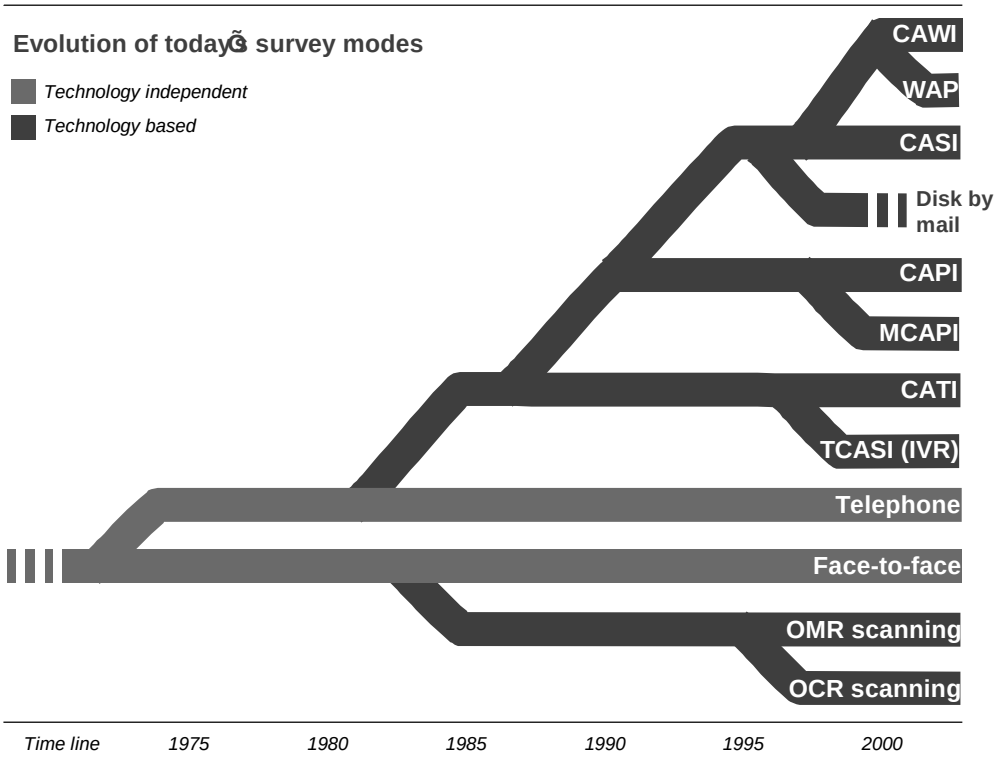


Figure 1. The evolution of today's survey modes

Almost all of the earlier interviewing modes survive to this day: with the possible exception of disk-by-mail interviewing, which the Web has largely rendered redundant. WAP and SMS could share the same fate unless it finds its niche as a specialised tool (Noyce et al 2003).

Just as few interviewing modes seem to die, research software packages are also remarkable for their longevity. In many cases, as additional interviewing modes have emerged, software manufacturers have been keen to implement these. This year's Research Software Review (Macer 2003) identified 42 different survey software solutions, classified according to six modes. Of these, 34 offered support for two or more modes, and 11 offered support for all six (see Table 1).

Table 1. Modes supported in survey data collection packages *Source Macer (2003)*

Total number of modes			**Cumulative number of modes**		
Modes	Products		Modes	Products	
2	5	12%	2+	34	81%
3	5	12%	3+	29	69%
4	4	10%	4+	24	57%
5	9	21%	5+	20	48%
6	11	26%	6+	11	26%

100% = total of 42 listed software products

Increasingly, manufacturers are claiming that their interviewing solutions are multi-modal and, either explicitly or implicitly, that this support extends to mixed mode research. After examining the issues raised by several research practitioners, this paper will consider the extent to which manufacturer's are understanding the needs of the mixed mode researcher and providing solutions that meet those needs.

2. The researcher's need for multi-modal surveys

In the past five or six years, in the United States, Web surveys have risen to become an almost undisputed replacement for paper-based mail surveys, switching one form of self-completion for another. The benefits and variations are relatively well understood, sufficient to satisfy researchers that the results are comparable and therefore useful (e.g. Owen 2001, Comely 2001).

Now, technology is finding a more important role than being simply a means to save time or reduce cost. Research is becoming more difficult and technology, particularly the sophisticated use of multiple survey modes, is looking like one of the few ways to increase both sufficient coverage to achieve effective sample incidence[1] and acceptable response[2].

Response rates are falling throughout the developed world. Poynter (2001) estimates that changes in society, including greater mobility and an increased awareness of privacy, now mean that response rates of below 20% are common, which is compounded by sampling methods that, at best, only reach 20% of the population, threatening all assumptions about coverage and response.

One proponent of mixed mode surveys is Don Dillman, who has published extensively on the subject since he proposed his Total Design Method in 1978 (Dillman 1978). At a Gallup symposium in the United States, he remarked:

> *"Survey organizations, whether they are in universities like mine, in private-sector organizations or in government organizations, are going to have to change dramatically in some ways in order to do effective surveys as we bring these new technologies online and still use our other technologies where they work."*
>
> *Don Dillman, in Quirk's Review (Crabtree 2000).*

It is a sentiment echoed by Craig and Douglas (2001), who now consider it imperative, in the context of any international research, that researchers should engage with a range of survey technologies in order to carry out research across international boundaries, particularly where these countries are at different stages of development or where there are disparities in economic and social conditions.

Case studies that demonstrate the efficacy of the mixed mode research are now starting to emerge. Allison and O'Konis (2002) demonstrated response improvements are achievable. In a survey of financial service customers in the United States, respondents contacted by telephone were given the option to continue the interview online: a staggering 88% chose this option, and 54% of them went on to complete the survey. Oosterveld and Willems (2003) have proved it is possible to design mixed mode surveys where mode has no influence on the answers given, in an experimental design which separated modal effect from population effects. They assert that the majority of modal differences reported in previous studies can be explained in terms of population differences. Allison and O'Konis also found high levels of similarity between responses elicited from the different modes. Dillman et al

[1] Incidence: the proportion of the target population actually reached by the chosen sampling method

[2] Response: the proportion of sample units completing a survey from any selected frame.

(2001), exploring the same issues across four modes have also concluded that, although there are observable differences, these are outweighed by the overall improvements in sample coverage and response, as well as time and cost.

Mixed mode research can be argued as being less burdensome on the respondent (Allison and O'Konis 2001 p 88), which not only provides better response for the survey in hand, but can be argued as becoming an essential consideration for the long term survival of opinion research as a viable activity.

The rationale for fielding a survey using a combination of modes is inevitably a trade-off between the benefits outlined above and the risks. In essence, the risks fall into three categories: calibration, coverage and complexity.

1. **Calibration** refers to the risk of differential measurement error due to modal effect on the respondent, where different modes yield different answers and would lead to different conclusions if each mode was treated in isolation.
2. **Coverage** relates to the large and well-aired subject of sampling, and in particular, the risk of differential non-response from the sub-samples contact through each mode.
3. **Complexity** refers to the risks of increased cost, delays and errors, arising from the duplication of effort or additional work inherent in conducting research across different modes.

Calibration and coverage issues need to be controlled through the research design, though technology has a role to play in providing appropriate support to implement these research decisions effectively. Issues of complexity can, and should be, managed through better, more effective technology.

3. Calibration and coverage: the issues

Presentational influences

It is well known that, from one survey to another, the style of the survey affects the response rate. Ganassali and Moscarola (2002) have quantified the effect on response of using a single scrolling screen or a single question per screen on response rate. They also determined that the presence of visual elements can produce a significantly higher response than the equivalent question with no visual support. This has implications for mixed mode research designs, if visual elements are provided in one mode and are absent in another.

Interviewer influences

The moderating effect of an interviewer in CATI and CAPI can give rise to under-reporting of personal or emotive issues. Poynter and Comely (2003 pp 175-176) described massive under-reporting on a question relating to the use of mobile telephones while driving in the UK, an emotive issue with anti-social connotations: on the Web 37% said they 'rarely' used a phone when driving, while only 21% admitted it on the telephone.

Differences in responses to open-ended questions

Oosterveld and Willems (2003), in a mixed mode research design, recorded a tendency reported by others for significantly longer and more detailed responses to openended questions recorded on the Web than on the telephone. However, Allison and O'Konis (2002 p 90) reported great similarity in terms of word length between the two modes.

Differences in responses to anchored scale questions

Humphrey Taylor (2000), noted a tendency for respondents on the Web to answer scales differently. Allison and O'Konis (2002 p 91-92) observed significantly more CATI respondents tending to use the extremes ('strongly agree' and 'strongly disagree') while CAWI respondents were favoured the middle ground. Dillman et al (2001) observed the same in their study and recommend great care in the use of classic scales with anchored points (e.g. "where 1 equals strongly agree and 5 equals strongly disagree") in mixed mode surveys. Yet the effect has been observed in mixed mode paper and Web surveys too, (Bäckström and Nilsson 2002) which means the distortion cannot be explained by differences between auditory and visual stimuli alone.

Hogg (2002 p 90) describes a tendency for respondents to select "don't know" and "no response" options more frequently in Web surveys, than interviewer-led CATI. His approach is to vary the survey instrument when delivered online to omit these in all but a few carefully judged situations.

Non-response error

One of the few experimental examinations of non-response error in mixed mode surveys is by Dillman and others (2001). Their observations do not offer encouragement to those hoping mixed mode surveys would provide high-tech relief to the growth of non-response in surveys. The study deployed four modes: CATI, CAWI, paper-based self-completion by mail and IVR-based TCASI and used a sample frame where demographic details were known. The characteristics of the non-response group were consistent across all four modes: younger, less well educated and, curiously, female; though they also observed the tendency for higher response from males on the Web and females on the telephone.

The above only presents a summary of the major influences. There are many different ways in which mode can influence not just the level of response achieved, but the actual responses given—and in ways we are still discovering. It is worth remembering that these effects are there in unimodal research too, but without the addition of the other modes, they remain latent. While we cannot look to technology to iron out the imperfections of modal influence, technology must support mixed mode in a way that does not unwittingly introduce more noise into the process.

4. Operational complexity

The increase in operational complexity is the most obvious and immediate effect of any decision to mixed modes in a survey design. It is felt most acutely by those responsible for administering and managing the survey. The problem areas to addressed can be characterised as follows:

Different recruitment and screening strategies

Unless all initial contact is to be made by the same mode, then the recruitment and screening method will vary from one mode to another, and different strategies will need to be devised then programmed into the survey instrument for each mode. This can be time-consuming to program and is logically complex, requiring careful testing to avoid programming errors.

Brown et al (2002) identify three 'hurdles and limitations' in using online research in a business-to-business context, as opposed to telephone research: the difficulty in achieving initial contact; the loss of any sense of dialogue, which is customary in both business and personal interaction and reliability from incorrect response to questions without the intermediation of the interviewer.

Duplication of the survey instrument

Many mixed mode researchers have been faced with a complete duplication of effort, when programming the same survey instrument into incompatible CATI and Web interviewing systems. A mixed mode survey system should be able to reduce this duplication of effort by allowing a single survey instrument to be executed in different modes, and allow for adjustments and variations to be made to ensure it executes appropriately in each mode.

In reality, much current survey technology falls well short of this ideal. Even where the same system supports several modes, to run two or more modes in parallel may still require different versions of the of the script for each mode, which is a potential source of error.

Data handling issues

A relational database, with its capabilities to allow multiple, simultaneous access to data, would seem to be the ideal means to handle survey data in mixed mode surveys, by acting as a common repository for all active and complete interviews. Achieving this ideal is difficult for numerous reasons. For a start, with their legacy of sequential files based solutions, very few interviewing systems actually support relational databases. Mixing offline and online interviews is also problem, despite developments in wireless technology, and the imposition of firewalls between internal systems and external Web servers can also make it difficult to use the same database throughout.

These issues were apparent in a large multimodal survey of 60,000 households in the Netherlands carried out by a consortium of research organisations, on CATI, CAPI and CAWI (Rippen 2003). Despite using the same software, it proved impossible to link the different databases, so the researchers had to build one from scratch, which was also not without problems. Rippen described the situation as follows:

> *"In practice, things turned out to be much more difficult to streamline… We had to create three different scripts for each interviewing method, and this made the database very large and unwieldy. It also required a lot more scripting hours than we imagined."*

Mode switching

Bäckström and Nilsson (2002) state the importance of providing respondents with the choice of method in increasing motivation to respond even though switching an interview from one mode to another can result in considerable operational complexity. Unless all of the modes share the same database, this also introduces delay as records are moved from one system to another. Apart from the scope for error, these delays can have an effect on the response. Switching an interview from CATI to CAWI, Allison and O'Konis (2002 p 88) observed a faster response once the software was adjusted so that, an email was despatched immediately the telephone portion ended.

Operationally, it was also a critical issue on the Dutch household study. Rippen (2003) observed that, without core support in the survey software being used, their own database only provided a *results* database, whereas what was needed to ensure timely and efficient switching between modes was a *fieldwork* database, driving each of the interviewing systems.

In addition to the four main issues described above, there are several other secondary issues that those designing or implementing surveys across modes have to consider, such as enhanced operational reports to able to obtain, clear across-the-board view of sample disposition, progress towards quota targets and the ability to define call-back strategies both within one mode and across modes.

5. What do software manufacturers mean by mixed mode support?

In the course of my research, I approached ten of the software manufacturers offering support for multi-mode surveys for information, with three research questions:

1. What did each consider to be the issues, constraints, operational and technical challenges of mixed mode research that survey software needs to address?
2. Which specific capabilities did each developer provide to overcome the operational challenges?
3. What technical solutions were each developer implementing or planning to implement to control modal influence?

Of the ten approached, eight went on to provide information for this research. The interviews were conducted by a combination of email, telephone and face-to-face. The responses are summarised in Table 2, with the number of mentions for each item shown in parentheses.

Awareness of the challenges

There was a wide variation between the responses from the manufacturers, with consensus on few issues. Several chose not to address the first question, preferring to answer in terms of what they were doing (the scope of the second and third questions). Suppliers chose to concentrate on operational complexity issues more than calibration and coverage issues. Coverage received no specific mentions.

Complexity issues tended to focus on the business of creating a mode-sensitive survey instrument from a single-source authoring tool, along with data handling issues and the consideration that a relational database is required to manage this. Though ease of modal switching featured universally in the second research question, it was rarely singled out as an up-front issue, (it may have been what suppliers meant in the three non-specific mentions of operational efficiency). Recruitment and screening received not a single mention.

6. The developer's response to calibration issues

This was the third research question, though consistent with the rest of this paper, it will be considered before complexity issues. Only two companies appeared to be making a serious attempt to understand and control modal influence. OpinionOne has developed a new hybrid interviewing method which it calls CAVI, for computer-assisted visual interviewing (and has trademarked the term). Among several innovations designed to eliminate modal influence, is the system's ability to generate a consistent appearance for each question and each screen across different hardware for CAPI, for CASI in kiosk interviewing situations and for CAWI. In CAWI, the Web browser expands to fill the whole screen, the characteristics of the PC's display are interrogated and determined and the size of the items on screen are individually recalibrated to ensure consistency. If the survey is translated into multiple languages, the positions and sizes of each button and text box is readjusted so that the screen layouts are identical across all the languages.

Another pioneer in this respect is Sphinx, which has undertaken a series of research experiments in order to examine the influence of mode, and of different research designs, on the data obtained (Ganassali and Moscarola 2002).

However, an important aspect of calibration is the ability to measure the response by the mode used to collect the data. Pulse Train goes further than any other manufacturer in its Bellview Fusion product, by recording the mode for each question, which takes into account the fact that an interview can

switch mode at any point in their system. While the data are recorded automatically, it is unfortunate that in the current implementation of Bellview Fusion, this paradata is not available to data record, and some complex and rather unsatisfactory manoeuvres are required to release it for comparison with the case data. Askia, Nebu, MI Pro and Sphinx also all record the mode of the interview in the data and a similar facility is believed to be part of the metadata within the SPSS MR Data Model.

OpinionOne has also developed a potential substitute for the unaided question online. In CATI or CAPI, it comprises a predefined and displayed list of items that the interviewer does not read out, but uses to record the respondent's spontaneous response. It uses a combination of an open-ended question and a capability to perform automated coding in real time against a user-defined dictionary.

Table 2. The challenges of mixed mode research from the producer's perspective

I. Calibration issues calibration between the modes (2) convincing users that the data obtained are heterogeneous (1) considerations when migrating continuous studies from a single mode to mixed modes (1) the need to establish a deep understanding of the inherent biases of one mode over another (1) reducing the burden on respondents (1) creating a bank of predefined, calibrated questions	**III. Complexity issues** operational efficiency: *non-specific* (3) (a) different recruitment/screening strategies (no specific mentions) (b) duplication of the survey instrument producing a single-source authoring tool for all modes (2) automatically vary the questionnaire by mode (2) definition of the modal context as well as the language for each question context specific validation, according to the mode identifying what systematic changes need to take place with a mode change separation of survey definition and execution process (2) finding an equivalent for unaided questions online (2) (c) data handling common database for all active interviews and results (3) development of appropriate XML schemas to ease data transfer common storage of sample management data single data management tool for all modes (d) mode switching in general (2) converting survey instruments from one system to another (e) other operational considerations cross-modal quota control (1)
II. Coverage issues (no specific mentions)	

7. The developer's response to complexity issues

Operational complexity appears to be area where manufacturers are putting in most effort. The modes supported for mixed mode research for each of the manufacturers researched is shown in Table 3, and shows that a basic level of functionality already exists in most solutions for most of the attributes.

Table 3: Overview of operational support for mixed mode surveys by manufacturer

	Askia	Mercator	MI Pro	Nebu	Opinion One	Pulse Train	Sphinx	SPSS MR
CATI, full	■		■	■		■		❖
CATI, light	■	■	■	■	❖	■	■	❖
CAPI	■	■	■	❍	■	❖	■	❖
CAWI	■	■	■	■	■	■	■	■
Paper	❖	■	■			❍	■	■
IIIa. Differential screening	■		■	■	■	■	■	■
IIIb.Single source authoring tool	■	■	■	■	■	■	■	
IIIb. Mode-specific templates	■		■	■	■	■	■	■
IIIb. Mode specific texts	■		■	■	■	■	■	■
IIIc. Central database	■		■	■	■	■	■	❖
IIId. Mode switching	❖	❖	■	■	■	■		■
IIIe. Live, cross-modal quotas	❖			■		■		■

Key ■ Fully supported ❖ Partly supported ❍ In development

The level of support has been determined by an assessment of the functionality offered within a single, consolidated application suite. For reasons of clarity, CATI is subdivided into *full* and *light*, according to whether the CATI system is appropriate for use in interviewing centres with 40 or more stations. Among the eight manufacturers listed, Pulse Train's Bellview Fusion and SPSS MR's Dimensions are in the process of making the transition from being legacy file-based systems to being re-engineered multi-modal database driven systems. Mercator's system remains sequential file based. The others were designed to be multi-modal and relational database-driven, from scratch or in an earlier version.

Design and modal independence

Most of these manufacturers have successfully separated the abstract design of the questions from mode-specific consideration of their appearance, by creating templates. In Askia, MI Pro, Bellview Fusion, Nebu (where they are called stylesheets) and SPSS MR Dimensions, a different template will be applied at the time the interview is executed, to adjust its appearance according to the mode. SPSS MR has usefully adopted the term "player" for a particular mode, which can take decisions on how to interpret the generalised survey script, both from reading the associated template and from its own innate knowledge of the characteristics of the current mode.

Nebu has sensibly provided one template for CATI, to ensure a consistent look and feel for telephone interviewers; the other templates refer to CAWI, and can vary by survey. In MI Pro, as you start to design your survey, you are confronted with a check-box list of the different survey modes. The system will automatically generate a version of the survey appropriate to each mode selected, determined by user-defined templates.

Central database

The central database is a characteristic of each of the software packages examined, with the exception of Snap. Askia, in particular, incorporates a lot of flexibility here. CATI and Web can either share the

same database, or if the Web server is to be outside the firewall, there is an interchange utility to exchange data automatically between the databases. Both Pulse Train and SPSS MR allow the Web, CATI and database server each to be deployed independently of the other, overcoming firewall issues, as the Web servers and the database server can be on opposite sides of the firewall.

Incorporating data from offline modes, such as CAPI or paper, can be time-consuming and error-prone. Wizards are provided in Askia to step users through the process of loading offline CAPI data or keyed-in data into the database. While Nebu provides support for GPRS (cellular data) connections from CAPI interviewing, with real time checking of quotas, even though the rest of the interview is not dependent upon a live connection being maintained.

Switching between modes

Modal switching is very capably handled in Askia, Bellview Fusion, Nebu and OpinionOne. Askia incorporates a special question type which allows you to select an alternative mode. According to the mode selected, the handover is then achieved automatically. In the case of a transfer CAWI, the email is generated and despatched automatically. In Bellview Fusion an interview can be 'abandoned' at any point in any mode and transferred to a different mode, either by scripting in a decision point as a question, or for interviewer administered modes, at any point on demand, by recourse to a special menu of actions.

A similar level of functionality is provided in Nebu though either its 'static switch' (a fixed decision point in the script) or 'dynamic switch' (can occur anywhere, on demand from interviewer or respondent), and a special question type called a 'dynamic swap object'. This is used to record vital next contact information, such as the respondent's email address. Another template determines the form and content of the email message which is then despatched automatically. Ingeniously, once a record has transferred to a self completion mode, the option to go backwards through the interview will not allow the respondent to go back into the CATI-collected portion.

Askia, Nebu and Pulse Train all allow for interviews to go back to CATI again. Pulse Train is extending the capability of its dialler software to allow provision of an instant 'call me back' button on Web interviews. It also allows for a conditional handover, by setting a CATI callback appointment at the same time as assigning the interview to CAWI. If the respondent fails to complete the interview on the Web, he or she will be contacted again through CATI.

The integration of print questionnaires in MI Pro makes it possible, if switching from CATI or CAWI to paper, for a scannable version of the survey instrument, that shows the responses to any questions already answered, to be printed and despatched to the respondent.

Missing features

Several of the researchers' concerns in mixed mode research were not convincingly addressed by the manufacturers and deserve attention:

- The need to cross-tabulate completed data by data collection mode, at a datum level (not merely a case level), as this is crucial in measuring modal influence.
- Support for the systematic removal of certain answers from certain modes, such as to vary "don't know" and "not stated" answers by mode.
- Greater consideration of the up-stream management of sample to improve coverage.
- Specific support to simplify the creation of parallel screening or sample eligibility identification strategies across different modes.

8. Conclusions and recommendations

Empirically, mixed mode research does appear to solve problems of sample coverage, time and cost, while delivering heterogeneous data—subject to several predictable and avoidable exceptions. Some suppliers of research software are now producing multimodal research packages that contain a range of technical innovations to facilitate mixed mode research. The focus of these developments is on overcoming operational complexity, but the other issues raised by nixed mode research, such as calibration and coverage, are largely being ignored.

On close inspection, there are relatively few survey packages that will support a range of basic needs in undertaking mixed mode surveys. The systems that do, including those examined here, tend to favour particular modal pairings: CATI/CAWI, CAPI/CAWI and so on. No single solution has yet emerged capable of handling the four-mode experimental study conducted by Dillman et al (2001) or even Rippen's (2003) three-mode study with a single authoring tool and a single results database.

For many developers, mixed mode research is still work-in-progress, and the range of modes supported and facilities provided will change rapidly from now. Developers need to remain focused on the needs of the mixed mode researcher and appreciate these go beyond just operational support. From the research undertaken, the core functionality to meet these needs is identified as follows:

1. A common survey authoring tool that generates a single survey instrument for all modes.
2. Independence between design and execution, with mode-specific templates and rules.
3. The ability to define mode-specific texts in addition to foreign language alternatives
4. A single, consolidated database for all survey data, updated in realtime.
5. Ability to determine the mode of initial contact from the sample subject's stated preference.
6. Efficient switching between modes, initiated by the script or by the respondent.
7. Ability to conceal all interviewer-recorded data when switching to self-completion modes.
8. Support for reminders and fall-back strategies to revert to a prior mode if still incomplete.
9. Single view management and reporting, which identifies response by mode.
10. Quota controls implemented across all modes in real time.
11. Question constructs for mixed mode, e.g. unprompted questions for CAWI, and the ability to have mode specific answer categories (e.g. "don't know and "not stated").
12. Recording of mode applied, at a datum level not just a case level.

Acknowledgements

The author thanks the following for their kind assistance in the research for this paper: Jérôme Sopocko at Askia, Steve Jenkins at Mercator, Per Jacobsen at MI Pro, Eric van Velzen at NEBU, JJ Klein at OpinionOne, Paul Quinn and Leonid Sokolo at Pulse Train, Jean Moscarola at Sphinx Development, and James Corriveau, Colin Linsky and Richard Scionti at SPSS MR.

References

Allison J and O'Konis C (2002) If Given the Choice, *Quirk's Marketing Research Review*, July/August issue, p 20.

Bäckström, C and Nilsson, C (2002) Mixed mode: Handling method differences between paper and web questionnaires, http://gathering.itm.mh.se/modsurvey/pdf/MixedMode-MethodDiff.pdf

Craig C S and Douglas S P (2001) Conducting International Market Research in the 21st Century, International Marketing Review, 18.1, pp 80-90.

Brown J Culkin, N and Fletcher J (2002) Human Factors in Business-to-Business Research over the Internet, *International Journal of the Market Research Society*, vol. 43.4.

Cobanoglu C, Warde W and Moreo, P J (2001) A Comparison of Mail, Fax and Web-based Survey Methods, *International Journal of the Market Research Society*, vol. 43, Quarter 4.

Comely P (2001) How to do Online Research, *Research Guide to Internet Technology*, The Market Research Society, London, UK

Crabtree S (2000) Untangling the Web, *Quirk's Marketing Research Review*, July www.quirks.com, quicklink 608.

Dillman D A (1978) Mail and Telephone Surveys: The Total Design Method, New York, Wiley

Dillman D A, Phelps G, Tortora R, Swift K, Kohrell J & Berck J (2001) Response Rate Measurement Differences in Mixed Mode Surveys Using Mail, Telephone, Interactive Voice Response and the Internet, AAPOR Annual Conference, Montreal, Canada

Ganassali S and Moscarola J (2002) Protocoles d'enquête et efficacité des sondages par Internet, Journées E-Marketing AFM/AIM Conference, Nantes, France

Macer, T (2003*) Research Software Review*, The Market Research Society, London.

Noyce D, Hofman O and Menti M (2003) Research on the Move. ESOMAR Technovate Conference, Cannes /www.skopos-mr.co.uk/market_research/skopos/studies/studies.html

Oosterveld, P and Williams P (2003) Two Modalities, One Answer. ESOMAR Technovate Conference, Cannes.

Owen R S (2002) A Matter of Trade-offs: Examining the advantages and disadvantages of online surveys, *Quirk.s Marketing Research Review*, February, pp 24-26.

Poynter R (2001) We've Got FiveYears, ASC Conference, London, September www.thefutureplace.com/five_years.htm

Poynter R and Comely P (2003) Beyond Online Panels. ESOMAR Technovate Conference.

Rippen H (2003) One project, three methods, in Macer (2003).

Taylor H (2000) Does Internet Research Work? Comparing online survey results with telephone survey, *International Journal of the Market Research Society*, 42.1

About the Author

Tim Macer is managing director of meaning ltd, an independent consultancy specialising in research, technology and organisational development. He is a specialist in the application of software and technology to survey research. Tim is a regular software reviewer for *Research* in the UK and *Quirk's Marketing Research Review* in the United States. He may be contacted at meaning ltd, Old School House, Main Road, Kingsley, Bordon GU35 9ND, telephone 01420 475700, or by email to tim.macer@meaning.uk.com

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Visiting Rural 'Virtual Villages' Through Email Surveys and Online Ethnography

Martyn Warren & Sarah Skerratt

Abstract

The proliferation of rural village websites in recent years, and websites which act as portals for these individual villages, can be seen as a demonstration of the perceived validation which such a web-presence gives. Images implying values of 'community' and 'neighbourhood' can be portrayed through this medium, with the aim, *inter alia*, of promoting general interest in the village and area, and providing services to its inhabitants. Perhaps more importantly, such 'virtual villages' appear to offer a medium for stimulating rural development, for instance by attracting tourists and businesses, by offering agencies access to ready-made communications networks, and by creating new 'communities', more or less grounded in the old, which can act as catalysts for initiative.

This paper examines the possibility for us as researchers to visit and understand these rural villages through virtual means, firstly, through an email survey of webmasters, and secondly, through 'virtual ethnography' (after (Hine 2001)). Firstly, we examine the types of data and insights that can be generated through the email survey, including the types of interactions which take place between webmaster and researcher, for example in terms of dialogue, offers of help, clarification and confidentiality issues. Secondly, we assess the potential afforded by this online technique, for the rapid validation of findings by respondents. Thirdly, we present and assess those data types which are accessible through an 'online ethnography' of the site masters and their sites, and the extent to which these are different from, and add value and depth to, the survey data. Finally, we evaluate the effectiveness of the email/online medium of data capture for generating findings about such place-based, online communities, and the extent to which this approach subsequently allows us to assess the potential for such rural village websites within rural development.

Keywords

Data capture; email survey; 'online ethnography'; rural; village websites.

1. Introduction

During 2002, the authors launched an investigation of village websites in the UK – the 'Virtual Villages' project. The thinking that led to this project, and its relationship to the literature, is described fully elsewhere (Skerratt, Warren *et al*. 2002). This paper, though, is concerned primarily with the *processes* used, and with the lessons learned from developing and implementing the methodology: for the same reason, it does not devote much space to detail of the findings of the project.

The broad aim of our research was to explore the implications of the development of village websites for the process of rural development in the UK. Given the nature of the topic, and of the likely target population, it made sense to consider the use of online survey instruments for at least part of the study: in addition, we were interested in exploring the potential for applying ethnographic methods through the medium of the internet.

2. Some methodological issues from the literature

Although the specific topic of rural community village websites analysis is not represented in the literature, there is a rapidly developing body of analysis which is relevant to development of methodology. Mann and Stewart supply an extensive appraisal of online methodologies, noting with surprise that the suitability of the Internet for conducting research remains relatively unexplored, with the focus more on what can be *observed* online (in terms of, for example, shifts in social networking), rather than how online approaches can be *used* for carrying out research (Mann and Stewart 2000: 7). Other authors highlight the importance of research focussing on the interactive communication abilities of the internet (Costigan 1999: xix) and draw attention to key advantages of the medium for researchers:

> *"Thanks to hardware and software, we have the artificial textual traces of interaction created instantaneously, at the moment of utterance. For researchers with an interest in discourse analysis … the Internet is a research setting par excellence…" (Jones 1999: 13).*

However, Jones cautions that the social issues of the internet "are far more difficult to untangle than its texts". He argues that one of the goals of computer mediated communication (CMC) research is what he terms "summative", that is, research "which aims to summarise how [a computer] system affected those involved with the system as well as the wider social context, including intended and unintended effects, and to what extent the system goals were achieved" (p.21). Further caution is advised by Sterne, suggesting a tendency for internet scholars to "universalise their own subjective impressions and dispositions, thereby grossly overestimating the impact, magnitude, accessibility, and universality of their object of study… Many writers have made wildly exaggerated claims about the ease of access to the medium, its relative importance…", and this has had direct methodological implications for online research (Sterne 1999: 278). Sterne therefore argues that "computers, access to networks and software literacy are themselves embedded in material and symbolic economies that require careful critical attention" (p.281).

How might such methodological and contextual complexities be addressed? Hine's text *Virtual Ethnography* is one example of a methodological exploration of how a pre-existing methodology (the practice of ethnography) can be applied to an online environment (Hine 2001). While acknowledging that there are conceptual complexities to address, since "face-to-face interaction, and the rhetoric of having travelled to a remote field site, have played a major part in the presentation of ethnographic descriptions of authentic" (p.10), Hine argues that such adaptation of methodology is possible, and desirable if one is to generate a rich understanding of internet as culture and as cultural artefact. Further, "conducting an ethnographic inquiry through the use of CMC opens up the possibility of gaining a reflexive understanding of what it is to be part of the Internet… An ethnography of, in and through the Internet can be conceived of as an adaptive and wholeheartedly partial approach which draws on connection rather than location in defining its object" (p.10). Hine advocates online participant observation on the grounds that this will allow the researcher to have similar experiences to those of informants (p.10).

A complementary approach is to utilise multiple methods, where ethnography sits alongside other approaches, allowing the researcher to attempt a triangulation of results (Mann and Stewart 2000: 67). Costigan cites a range of disciplines from which these can be drawn, including communication research, media studies, anthropology, sociology, literary criticism, cultural studies, psychology, and political economy. Questions remain: "Which are useful? Which are not? Which do we believe will contribute most to our knowledge of the Internet?" (Costigan 1999: xi-xii). Similarly Sterne argues for a hybrid approach, combining, in various degrees, ethnography, autobiography, and textual analysis (Sterne 1999: 269).

Boundaries between offline and online worlds are also debated in the literature: i.e. researcher-imposed, or prescribed, data collection boundaries (which may reflect implicit conceptual boundaries). Various researchers (e.g. Kendall 1999; Miller and Slater 2001) attempt to overcome this by use of a mix of offline and online research methods, for instance on-line participant observation, face-to-face interviews, participation in face-to-face gatherings, and reading of related newsgroups and email lists (Kendall 1999: 58). Kendall justifies his approach thus:

> *"Once on-line, participants draw on their off-line resources, as well as understandings gained in off-line experiences, to negotiate and interpret their on-line interaction… researchers need to consider participants' local off-line environments, as well as to explore how participants blend their on-line and off-line lives and social contacts. Individuals exist and participate in off-line social contexts both sequentially and simultaneously with their on-line participation. However, many accounts of on-line spaces, experiences, and communications ignore this, often describing on-line spaces almost wistfully as a new and discrete utopian world." (pp.58 and 60)*

The fact that our research relates to community adds a further complexity. Work by Harrison and Stephen is helpful in outlining some key issues when investigating online community, highlighting a key methodological consideration: that of reflecting on how we conceptualise, and therefore 'frame', research into community networks:

> *"Most researchers regard the Internet and the World Wide Web as technologies that are transforming the world into a 'global village'… In contrast, a rapidly growing community networking movement has chosen to use the Internet and the WWW as a resource to enhance the development of geographically based communities" (Harrison and Stephen 1999: 221).*

This again emphasises our need for an array of methodological tools which allow us to investigate the phenomenon of rural village websites, and iteratively to select and combine those approaches which generate the clearest understanding of this phenomenon. This 'toolkit' evolves alongside our understanding. To quote Hine: "As an analytical starting point it is unproductive to take any features of the technology as taken for granted and simply the way things are…" (Hine 2001: 8)

Such observations, from both methodological and conceptual standpoints, are critical to our analysis of what we have termed 'virtual villages', and have significantly influenced our research design.

3. Research design

We initially identified six phases, which have become further subdivided during the course of the project: at the time of writing (April 2003) we have completed phases 1, 2a, 3, 4a and elements of 6. Phase 4b is in preparation. The intention was to identify and classify the population of village websites, using this information to survey a sample using online approaches. This would be followed by a 'funnelling-down' approach, using online methods to probe issues in more depth with a smaller group, and finally selecting a small number of case studies for closer offline study, ideally using

ethnographic techniques, in order to compare and contrast online and offline 'communities' (objective 5). Some element of triangulation would be provided by participant validation of findings at each stage, and through in-person interviews of rural development specialists working extensively with online approaches. A recent decision has added the creation of an online discussion group or forum enabling us to deploy some elements of online ethnography. The overall design of the research process is shown in the following diagram:

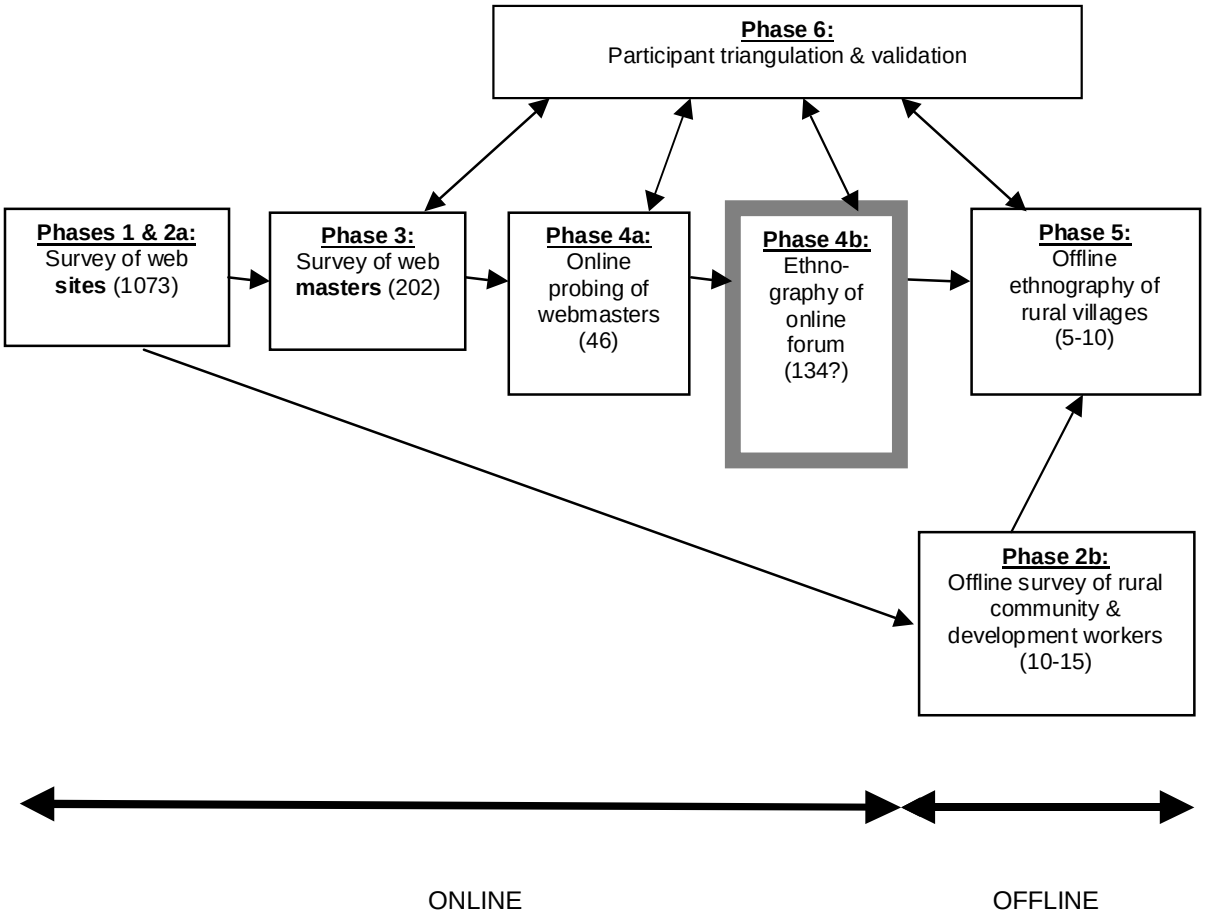


Figure 1: Phases of research design

Phase 1: search for rural village community websites

For the purpose of this investigation, we defined a 'rural village community' website as having the following characteristics:

- The village/community is situated in a rural location, and
- The site has been developed by community member(s), for the community.

The websites were located through systematic use of internet search on the key words 'UK', 'rural', 'village' and 'community'. A snowballing technique was employed to locate additional sites using links from web portals/rings 'Villages Online' and 'Irish and British Villages', and from other sites of interest. This produced a total of 1073 UK 'village' websites from which an initial 40 sites were randomly sampled for a pilot study of preliminary reading and analysis. A further sample of 469 'village' sites was subsequently drawn randomly in order to provide 202 sites for analysis (approximately 20% of the identified population). Spurious data were discounted (146 websites in

total) but were categorised to 'type' in order to inform decisions concerning the later development of a typology of 'village' websites.

Phase 2a: development and use of a protocol

A preliminary content analysis of 'rural community village websites' facilitated the identification of a number of recurrent themes, issues and characteristics, and the subsequent development of a 'protocol' or data collection sheet, which was tested and revised in arriving at the final version, together with a coding scheme for analysis. Data were gathered using both quantitative and qualitative content analysis techniques, which involved coding data from each site and taking notes while coding, thus enabling attention to be given to themes/issues that were of particular interest (such as services, breadth and depth of content, degree of physical rootedness, etc). For each website, the village name, location and URL details were recorded, together with contact details of the webmaster (for Phase 3). Content analysis identified key interests, themes and other characteristics, which as well as informing design of survey instruments, provides a 'stockpile' of information which will be used at a later date to cross-reference with survey and case study responses.

Another output of this stage was a typology of websites. This developed from an initial classification of sites according to the 'image' they appeared to convey, involving 13 categories. These were distilled further into a more manageable six categories:

- Ancestral (mainly concerned with genealogy) (9% sample)
- Strong community (online and/or offline: comprehensive coverage of community information and services) (24%)
- 'Select few' (online and/or offline: by and for a particular clique or section of the community) (18%)
- Marketing (primarily of the village as tourism destination) (16%)
- Hybrid (a combination of two or more of the above) (14%)
- Other (diverse, and difficult to classify with precision) (19%)

Any such typology is vulnerable to criticism of the choice of underlying criteria: they depend on subjective judgement, and a different judgement could result in a very different typology. However it provides a useful framework on which to build a picture of website provision, and an aid to purposive sample selection (in Phases 4a and 5).

Our approach has similarities to that used by Harrison and Stephen, in their work on community networks, although they answered all their survey questions through consulting only the web pages, and they describe their data interpretation as "quite liberal" (Harrison and Stephen 1999: 239), whereas we aimed to develop a protocol and coding system which could be the subject of subsequent analysis itself. This is important for us since, in addition, we wish to present our online website analysis to the webmasters offline (in Phase 5), in order to allow us to pursue any apparent discontinuity between the web image and the intended message, as well between the web image and the offline environment. We take note of lessons from research into personal web pages, from which:

> *"… it appears that there is a contradiction between the [webmasters'] perceptions, as reported in the survey, and their behaviour when constructing their homepages … it is not always the medium that creates the message… This possible tension between attitudes and behaviour suggests a direction for further research. A new dependent variable can be generated by content analyzing the homepages." (Gibbs, Matie et al. 1997: 21).*

Phase 2b: offline survey of rural community & development workers

This phase is yet to be completed. A purposive sample will be drawn of individuals actively involved with e-delivery of services, encouraging communities to 'get on-line', and e-Government initiatives, together with individuals involved in community or rural development and who also have ICT priorities within their remit. They will be interviewed in order to explore the implications of the development of rural village websites for social inclusion/exclusion in rural settlements; and to test perceptions of the potential roles and values of village websites in the development of the rural economy, society, culture and environment, through generating broader principles and patterns.

Phase 3: online survey of web masters

A structured questionnaire (with some open-ended questions) was administered via email to the sample of 202 village webmasters (the term 'webmaster' is not used here in a gender-specific way). We used the questionnaire to address each of our research objectives, with questions concerning the respondent (e.g. age, involvement in the village/community); the origins of the site (e.g. age, purpose, funding); current operation (e.g. who runs it, their motivations); the current nature of the site (e.g. intended and actual audience, resourcing); the perceived value of the site to the community; and future development.

When administering an online survey, the issue of sampling frame, and who one might exclude through using an exclusively online medium, is a key consideration. However, since we were interviewing those who present themselves online, we felt that we would not be introducing bias into our sample by using online interviewing. The issue of potential bias *within* an online sample, towards those who are more able to use the text medium, is addressed in Phase 4a (below).

The questionnaire was administered using *Perseus Survey Solutions* software. This allows rapid construction of questionnaires using a variety of question types (including open-ended) and their conversion into an html file for emailing as an attachment. On completion of the questionnaire, the respondent hits a 'submit' button in order to send an email with encoded results to a designated recipient account. Data is easily retrieved *en masse* from these emails, and simple analysis (e.g. frequency analysis and charting) performed within the Perseus software. Data is easily exported to other software such as Excel (quantitative and categorical data) and word-processing or textual analysis packages (textual responses). The selection of the software was itself an important component of the research process, since its user-friendliness is a key facet not only of overall response rate, but also of quality of responses to individual questions.

"The way interviewers gain access to potential participants, and make contact with them, can affect every subsequent step in the interview process" (Mann and Stewart 2000: 82). Great care was also taken over the introductory email message, the visual impact of the questionnaire, the sequencing of questions, and the relevance of question content, in order to encourage high response rates. We considered an important aspect to be the use of warm, welcoming and non-technical language – people who use technology are not necessarily technical people, and certainly most of our respondents, despite the grandiloquent appellation of 'webmaster', were self-taught rather than highly-trained. Another key element was a clear statement about why we were doing the work, how it might benefit those running village websites, together with a promise of feedback on the results (see *Evolving a set of guidelines for ethical online research*, particularly Guideline 2, (Sharf 1999: 253-255)).

A pilot survey of 19 respondents had an encouraging response of 13. Respondents were asked for suggestions for improvements, resulting in a number of additional questions, and the use of rating

scales instead of yes/no responses for others. The result was a 53-question questionnaire sent out on 10 October 2002. 55 responses were received within 3 days and 100 within 16 days: two reminders (at 13 days and 27 days) produced small flushes of responses, and the last response was received at 72 days after initiation[1]. The total number of usable responses was 134. This was 66% of the original sample, but after allowance for unobtainable webmasters (inoperable email addresses and no response to repeated telephone calls), amounts to a response rate of 75%. One respondent immediately questioned whether the message was genuine (partly because it had been transmitted via the Perseus server). This prompted us to send a follow-up message to confirm the status of the message. This was a fortuitous step, as it later transpired that it had encouraged several waverers to respond, and will be planned into future surveys.

Another frequently-expressed concern was our use of an html attachment for the questionnaire – several respondents stated that they never opened attachments (for fear of computer viruses, which were much in the news at the time). We responded by sending a text-based version embedded in an email – slightly less convenient and attractive, but compatible with the Perseus system. Reminder emails included the text-based version as well as html. The result was a rich set of quantitative and descriptive data, from which a summary could very quickly be drawn, to be posted on the University of Plymouth website (http://www.sh.plym.ac.uk/vvsurveyfeedback.pdf). This is not the place to present results, but it is worth noting that 79% of the village websites surveyed were operated by single individuals, using their own resources, motivated largely by a desire to 'do their bit' for the community, and with little if any contact with other webmasters. Respondents were invited to comment on the results as a form of participant validation: several took up the opportunity: most to offer thanks for doing the work, many to express interest in learning that others shared their own circumstances and problems, and to ask for some way of making contact. This latter point was key in our later decision to create an online forum as Phase 4b, which was also influenced by the way many respondents initiated and maintained ongoing dialogue with us through the survey phases – very different from other investigational techniques we have used (e.g. postal, telephone, face-to-face).

We were encouraged by our high response rate for our online survey, particularly since some authors predict low levels of participation for internet-based surveys: "[A]lthough a 50% response rate is typically considered minimally adequate for much traditional survey research, response rates around 20% are not uncommon for unsolicited surveys, and response rates to on-line surveys may be 10% or lower" (Witmer, Colman *et al.* 1999: 147). The reliability of this as a benchmark is called into question, however, by (unusually frank) comments of the same authors on their own survey process: "Firstly, the questionnaires appeared fragmented because there was no internal coherence between stimulus items. Second, the end users probably had low interest in the survey topic, which was unrelated to their newsgroup themes. Third, our relatively bland subject matter, coupled with the

[1] One of the key strengths of online survey techniques is their asynchronous nature, particularly compared with face-to-face or telephone interviewing, where agreed appointments have to be arranged with the former, or often immediate responses are required for the latter. The methodological literature points to the advantages of asynchronous communication, (for example: "CMC is user-friendly in terms of making rapid connections between individuals in an environment of their own choosing… unlike the phone, electronic communication need to not be intrusive or peremptory. With asynchronous forms, messages wait until the receiver is ready to attend to them"; Mann & Stewart, 2000, p.24) and we found that responses were spread fairly evenly throughout the day, with extremes 05.47 and 22.36. This attribute of online flexibility may be particularly pertinent when dealing with volunteers who may be running these websites outside of usual working hours, and can therefore choose a response opportunity that is feasible for them.

repetitious pattern of unrelated statements, may have appeared trivial or boring to many of our questionnaire recipients" (pp. 155-156).

Phase 4a: further online probing

Following our original research design, we used the analysis of the survey results, and in particular the thematic analysis of the open questions, to identify key issues to probe in greater depth, though still online. Once again Perseus software was used. Seven statements were constructed, a mixture of positive and negative (e.g. 'Development of the internet generally, and my village website in particular, has the potential to make a real difference in the way my village community functions', and 'Encouraging the use of the internet for shopping, accessing government services, online banking, etc, will add a further and unwelcome threat to the supply of services in rural areas'). Respondents were asked to register their reactions to each statement on a five-point scale between 'strongly agree' and 'strongly disagree', and then to elaborate on the reasons for their response. The questionnaire was sent to a sub-sample of 46 of the respondents to Phase 3 (just over one third), purposively selected. The number surveyed party reflected practical concerns, in that we needed to keep data generation within our capacity to collate and analyse. It also reflected our subjective judgement, based on the first survey results, that in a population with relatively little internal variation, this should be enough to give a reliable reflection of the range of views. The primary criteria for selection of the sample were the disposition to give reflective responses to the open questions in the first survey (by reference to responses to two of the more demanding questions) and the speed of response to the first survey. We reasoned that this would ensure the greatest magnitude, quality and speed of response, while recognising that there might be a bias against those who find typing and/or expressing themselves in print to be difficult: a small risk given that we are sampling a population who put themselves online and choose to communicate online through their website, and with visitors to that website. Secondary criteria included age and gender of the webmaster, size of the village, and whether the site was stand-alone or one of a group managed by the one person. Since analysis of the first survey had shown no significant association between any of these variables and the responses given, it was felt that it was sufficient to perform a visual check on frequency distributions to confirm adequate representation.

Of the 46 questionnaires sent out, 33 were returned (72%), supplying a large volume of rich data. Thematic analysis, using *HyperResearch*, is still ongoing, but major themes emerging include the crucial importance of local content; the equation of 'outside' influences/drivers with interference and loss of local relevance; the tussle between website about the village, or website for the village; and the internet as a positive asset in rural locales.

Phase 4b – 'online ethnography'

One of our hopes was that we could explore the use of 'online ethnography' or 'virtual ethnography' through the email survey and subsequent probing. On reflection, we considered that these processes did not justify the use of the term, since we had not been observing, and participating in, a naturalistic, online environment (see, for example, Sharf 1999; Hine 2001: 10). However, our experiences during Phases 3 and 4a gave rise to the idea of creating an online forum with our respondents as members. While giving them the opportunity of making contact with each other and sharing experiences (which some survey participants had indeed requested), it would also give us potential for an element of participant observation in a context where the content and direction of the discussion is determined by the members with minimal intervention from ourselves. Thus the intention is to create a discussion group, probably through *Yahoo Groups*, with (in the first instance) all the respondents to the Phase 3

survey as members. We have no idea whether this group is an appropriate size, but given the alacrity of its members to participate in surveys, we feel we can expect a high level of engagement. In creating the group, a few key issues will be 'seeded' in order to provide a stimulus to discussion: our hope is that thereafter the discussion will take on a life of its own.

An issue we will need to resolve is whether to be prepared to facilitate actively after the first week or so in order to keep the discussion alive, or simply to let it run its natural course even if this is fatal to the forum (see Mann and Stewart 2000: 156-157). The latter would be unsatisfactory in terms of production of data, and of utility of the forum to the webmasters, but the observation and recording of the process could nevertheless provide useful lessons for development of research methodology. An alternative to facilitation would be to widen the potential membership of the forum to, say, all identified webmasters. Other difficulties we might face include the loss of control/focus in the discussion in relation to our research objectives; the possibility of inflexible software limiting our options; and the analysis of a large volume of disparate data. However, given that we have already progressed as far as we wish with a highly-focussed online approach (Phases 3 and 4a), the risks appear manageable. Against them, we have many potential benefits. By using a different online approach from before, we will be able to see how people discuss issues amongst themselves – what terms are used, what is taken for granted and not discussed, and what is disputed. Thus, the *diversity* of situations, objectives, justifications etc, within our sample should come through. This diversity will further allow for a diversity of respondent validation – itself very valuable, enabling us to see and assess the heterogeneity of the group. A different approach might be more 'refreshing' to respondents (who will have been through two survey phases) and therefore might maintain willingness to participate. If the group discussion moves on from our initial pointers, then this may then yield data that we had not anticipated and that would not have been forthcoming in a context where we initiate and lead.

In writing about online focus groups, Mann and Stewart suggest that:

> *"…the main feature of the focus group method is that it is 'interaction focused'. It is this interaction which is thought to illuminate what people feel or think, as well as why they may feel or think in this particular way". (Mann and Stewart 2000: 99, emphasis in original)*

We aim to go further than the focus group, by creating an interactional space for village webmasters, within which individuals may seek to explain, correct, and enlighten one another, possibly with specific reference to their geographical, place-based locations and experiences, and with little directional involvement of the researcher. We expect thus to generate richer dialogue than is possible through one-to-one dialogue.

Phase 5: offline ethnography of rural villages

A vital part of our study will be case study research on a small number (6 to 10) of the villages represented in the Phase 4a sample. The sampling will be purposive, rather than aiming at being statistically representative, and will be illustrative of 'types' of 'process and product' of rural village community websites. Webmasters will be asked if they would be willing for an offline ethnography of the village illustrated by their rural community village website. In addition to gaining their permission, we will engage in a process of discussion over access to the village with other members of the village community/communities. This phase of the research will provide us with an invaluable comparative data set, and will allow us to address, particularly, Objectives 4, 5 & 6. We anticipate the sample size being approximately 6 to 10, although - as with Phase 4 – this will be determined by a combination of participants' willingness, and our own methodological considerations. Once again, we will be

gathering data relating to the original web survey and the broader theoretical research issues outlined above, as well as participants' unprompted information/data. In addition, this offline Phase will give us the unique opportunity to examine offline with online, to carry out ethnographic analysis of social and informational networks, and to assess the extent to which the internet (through the village website, for example) is 'embedded' in such communities. That is, we will be able to contextualise our earlier findings. The analysis will be thematic, and through the use of *HyperResearch*, a multimedia data set will be analysed and presented.

Phase 6 – Participant validation and triangulation

This Phase runs parallel to Phases 2b, 4 & 5. We view it as a key element in assessing, and re-assessing, the validity of our findings, and are therefore creating opportunities - both offline and online – for participant checking of our data and analyses. At its simplest, this has involved recording spontaneous feedback given by respondents (for instance in response to our efforts to set up a discussion group). More systematic was the specific invitation to respondents to comment on the summary results sent to them after the first survey: this will be followed by a similar opportunity to feedback on the second survey. Offline interviews will give opportunity for triangulation by cross-checking interview responses against survey responses and our analysis of the websites. These data will subsequently be used to qualify and enhance the data from the other Phases.

4. Conclusions

Our aim in writing this paper is to share the lessons learnt, and the processes we have gone through, in devising a survey methodology based on use of internet technology. The research project was not specifically designed as a methodological investigation *per se* (unlike, for instance that described by Mann and Stewart (2000)). However, as the research has evolved, we have been actively reflecting (through many iterations) on the data types being generated through the different Phases, and hope that our paper, in attempting to crystallise those reflections, will be of value to other practitioners.

The growing body of literature on internet research was used to identify general online research issues, and to inform our research design. In particular, we took care to incorporate multiple methods; to ensure a combination of online and offline; and to learn from cultural studies (using what Sterne (1999) might call a 'hybrid approach'). Our work relates well to community networks literature, focussing on a range of rural community village websites, and developing a protocol and typology which assesses the extent to which sites are 'serving' the community.

Most important, in the context of this conference, is the impact of technology on our survey processes. Firstly, it is important to acknowledge the enormous benefit we gain from using totally mundane computer software. Without, for instance, the use of word processor and spreadsheet, production of the protocols for 202 websites and their distillation into readily usable data would have been impossible given the small resources we had available. Although we tend now to take such facilities for granted, it is important to remind ourselves of the enormous impact they have had on the capability of researchers. A little less commonplace is the analysis software we used: SPSS for the quantitative data, allowing rapid and flexible interrogation, and HyperResearch giving the facility for basic but invaluable textual analysis. Again, such software enables a small and under-resourced team to tackle significant projects, and imparts a flexibility and speed of reaction that is invaluable in researching a fast-moving phenomenon.

Most important of all, however, has been the use of email surveying techniques. They can, of course, be used only for populations where it can be reasonably assumed that use of email is ubiquitous. In our research we could take it as a given, but there will be many other circumstances (research amongst academics, government officers, computer specialists, etc) where this would be appropriate. Using email as a vehicle meant that our survey processes were rapidly implemented, flexible and – important for a non-funded project, cheap – the marginal cost of each communication to the team was zero. Adjustments could be made rapidly (for instance the follow-up email in Phase 3, reassuring respondents that the survey was genuine); feedback of results to respondents was quick and easy, and in many cases an ongoing dialogue developed naturally with individual respondents, giving further insight into their perceptions and experiences. These characteristics, coupled with the informality of expression encouraged by the medium, have no doubt had a significant impact on the extremely high response rates we achieved, and have helped us to move quickly through phases of the research, maintaining momentum and currency.

Crucial to this process was the use of specialist online surveying software, Perseus Survey solutions. For a relatively small initial cost (around $650), this gave us facilities for a variety of email and web-based survey techniques, of which we used only a small part, together with basic analysis and presentation facilities. This software was a delight to use: question wizards made the compilation of questionnaires very easy; the resulting html files were attractive in presentation and easy for respondents to use; and the collation of incoming data was simplicity in itself. Export of data to MS Excel was straightforward: the biggest limitation was the lack of direct export as SPSS files (available in more costly versions), although this could be achieved via Excel with some small loss of functionality. Without this technology it is doubtful whether we would have been able to launch this investigation; it is certain that, if we had, we would still be struggling with analysis of the first survey.

At the time of writing, it is not possible to discuss the impact of the use of discussion group or chat room software, although this has become a key element in our design. Managing this effectively is one of our biggest remaining challenges, and we look forward to being able to report on our successes and failures during the conference itself.

References

Costigan, J.T. (1999). Introduction: Forests, Trees, and Internet Research. *Doing Internet Research: Critical Issues and Methods for Examining the Net.* Jones, S. London, Sage**:** xvii-xxiv.

Gibbs, J., Matie, S., Mandavil, A. and Yi, H. (1997). Expression of local/global identity in personal homepages. *Rochester Institute for Technology Communication, Technology and Cultural Values*, Rochester, NY. http://matei.org/natid/natidreport.rtf

Harrison, T.M. and Stephen, T. (1999). Researching and Creating Community Networks. *Doing Internet Research: Critical Issues and Methods for Examining the Net*. Jones, S. London, Sage**:** 221-241.

Hine, C. (2001). *Virtual Ethnography*. London, Sage.

Jones, S. (1999). Studying the Net: Intricacies and Issues. *Doing Internet Research: Critical Issues and Methods for Examining the Net*. Jones, S. London, Sage**:** 1-27.

Kendall, L. (1999). Recontextualising "Cyberspace": Methodological Considerations for On-line Research. *Doing Internet Research: Critical Issues and Methods for Examining the Net.* Jones, S. London, Sage**:** 57-74.

Mann, C and Stewart, F. (2000). Internet Communication and Qualitative Research: A Handbook for Researching Online. London, Sage.

Miller, D. and Slater, D. (2001). *The Internet: An Ethnographic Approach*. Oxford, Berg.

Sharf, B.F. (1999). Beyond Netiquette: The Ethics of Doing Naturalistic Discourse Research on the Internet. *Doing Internet Research: Critical Issues and Methods for Examining the Net*. Jones, S. London, Sage**:** 243-256.

Skerratt, S., Warren, M. and Cousins, K. (2002). The virtual village: bridging or widening the digital divide? *Cities & Regions in the 21st Century*, Newcastle-upon-Tyne, UK. September 2002

Sterne, J. (1999). Thinking the Internet: Cultural Studies Versus the Millennium. *Doing Internet Research: Critical Issues and Methods for Examining the Net*. Jones, S. London, Sage**:** 257-287.

Witmer, D.F., Colman, R.W. and Katzman, S.L. (1999). From Paper-and-Pencil to Screen-and-Keyboard: Towards a Methodology for Survey Research on the Internet. *Doing Internet Research: Critical Issues and Methods for Examining the Net.* Jones, S. London, Sage**:** 145-161.

About the Authors

Warren and Skerratt are equal authors. Martyn Warren is Head of Land Use and Rural Management, University of Plymouth: mwarren@plym.ac.uk. Dr Sarah Skerratt is Senior Research Associate in Rural Social Informatics, University of Newcastle upon Tyne: S.J.Skerratt@ncl.ac.uk. They share a research interest in ICT in rural communities, encompassing rural broadband internet, adoption of internet by farm and other businesses, as well as the village website project featured here.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Internet-Based Stated Preference Surveys. Studies in Traffic Information and Public Transport

Marit Killi & Åse Nossum

Abstract

Data collection by traditional computerized methods, like home interviews using portable computer, has become very expensive. During 2002 the Institute of Transport Economics in Norway carried out two Stated Preference (SP) surveys using Internet. This has proved to be less expensive and have, in general, the same advantages as other computerized interviewing methods. Using computers in Stated Preference surveys provides a great advantage in the quality of the data collections compared with manual methods. Computer based surveys allow the design to be customized to each respondent and hence improve data quality. To get a representative sample on Internet, one must be very careful choosing recruiting method. This paper presents experience from the two SP-studies carried out in 2002, and focuses on how to get a representative sample and how the recruiting method affects the response.

Keywords

Internet survey; Stated Preference; representative sample; recruiting

1. Introduction

Data collection by traditional computerized methods, as home interviews using portable computer, has become very expensive. During 2002 the Institute of Transport Economics in Norway (TOI) carried out two Stated Preference (SP) surveys using Internet software. SP-survey is a collective term for "questionnaires with hypothetical questions". Using Internet software has proved to be less expensive and have, in general, the same advantages as other computerized interviewing methods. Using computers in Stated Preference surveys provides a great advantage in the quality of the data collections compared with manual methods. Computer based surveys allow the design to be customized to each respondent and hence improved data quality. It makes it possible to produce customized questions for each respondent, in the sense that figures presented in some questions could depend on answers from previous questions. That is, for instance, the data the respondent reveals about her regular trip to work could be used to generate questions concerning such a trip to work.

Based on the two main Internet-based surveys during 2002 at TOI we will look at experience and draw conclusions about using Internet for similar surveys. The two surveys are "*Travellers' valuation of traffic information based on work journey by car*", funded by the Public Roads Administration in Norway and "*Public transport in the Oslo region -Travellers valuation of time*" funded by AS Oslo Sporveier et al. The two studies are similar in design and length.

The first study looks at drivers' valuation of reduced travel time variability and delay, new forms of information in order to reduce travel time variability and delay, and the value of such information.

The second study looks at passengers in public transport and their valuation of time saving. The motivation was to find indicators of passengers' preferences for alternative quality improvements, and to compare the findings with a similar survey that was carried out in 1992.

A more thorough discussion of the two studies will be carried out in section 4 and 5. Section 2 describes the design of the two surveys, while we take a closer look at the arguments for using Internet in section 3. In section 6 conclusions are drawn.

2. Survey design

When we decided to carry out the surveys on Internet, we chose the Sawtooth Software. Sawtooth Software has a module designed for programming traditionally questionnaires for Internet and one relative flexible Stated Preference modules. By using Sawtooth Software, we developed Internet surveys using easy-to-use templates, by pointing, clicking, and typing information into text fields. The surveys included skip patterns and randomisation of questions. We used response verification to ensure that certain questions were answered. We also used the possibility to show answers to previous questions on the screen. The most important reason for choosing the Sawtooth Software was however the fact that it makes it possible to use Stated Preference analysis for Web, which was crucial for our study.

When designing the questionnaire we discovered that our ideal questionnaire was more complicated than the software usually handles, e.g. when calculating travel time in minutes by subtracting arrival time from departure time, i.e. 1 hour equals 60 minutes not 100 minutes. When we tried to identify traveller's willingness to pay for reduced (or increased) congestion time another problem arised. In these questions, the new value (in terms of minutes) was calculated by subtracting ten percentage points from the percentage of travel time that took place in congested traffic. Likewise, increased congestion was calculated as a ten-percentage points increase. If the initial percentage were below ten, the new value would be zero. If the initial percentage was above ninety, the new congestion time became equal to travel time. This to insure that the congestion time did not exceed travel time. To our knowledge this is too complicated for the software to handle without any additional programming. Using additional Java Script programming in combination with Sawtooth Software solved these problems.

Stated Preference techniques are commonly used in valuation studies in transport and were chosen in both surveys for several reasons. With SP the participants of the survey face hypothetical choice situations. As opposed to revealed preference techniques, where the researcher observes the participants' actual choices, SP can have many hypothetical choice situations per respondent. Hence SP requires fewer participants. Another advantage of SP is the possibility to present choices that are not yet introduced in the market, such as a more advanced traffic information service than the service available today.

Two kinds of SP techniques were used in our surveys: *Contingent Valuation* and *Stated Choice*.

Contingent Valuation

Contingent Valuation, a method similar to Transfer Price, is direct valuation of some of the factors by specification of price. In the survey about travellers' valuation of traffic information we were interested in different attributes of a trip to work.

The values of each factor (here travel cost €5.70, travel time 45 minutes, congested part 18 or 14 minutes) were based on the respondent's answers to previous questions.

The attributes of a trip to work in this part of the questionnaire was: travel cost, travel time, arrival time variation, congestion time, and type of traffic information.

Each question included travel cost in combination with one other attribute, except for one question were the respondent was asked about willingness to pay for several simultaneously improvements.

In the public transport survey there were four questions like this. The respondents were asked about their willingness to pay for one improvement at a time. The questions included passengers' direct valuation of reduced travel time, reduced headway, sitting instead of standing and direct connection instead of transfer. The levels of the factors were based on previous answers given in the questionnaire. The cost of a public transport journey was included in all four questions. An example of this type of question (in translation) is shown in Figure 1. Prices were originally in Norwegian kroner (€1 = 7.73 NOK).

The cost of your journey is €2.80 and you need to change bus once.
Imagine the same trip without transfer.
What is the maximum amount you would be willing to pay for this direct journey?
That is, at which cost are the two journeys of equal worth to you?
Journey 1: Changing buses and €2.80
Journey 2: No transfer and €_____

Figure 1. Example of a question in the Contingent Valuation section. Public transport study

Stated Choice

Stated Choice is a choice between two journeys where several factors vary systematically at the same time. In each choice situation, two trips A and B with three or four attributes were presented on the screen. The respondents could indicate which of the two trips they preferred by ticking either "Definitely A", "Probably A", "Don't know", "Probably B" or "Definitely B". From such a sequence of choices it is possible to estimate relative valuation for the different trip attributes. An example of a choice situation from the traffic information survey is shown in Figure 2.

The traffic information survey included three choice sequences with nine choice situations in each sequence. The two factors, cost and travel time were included in all sequences. In addition there was a third factor that was specific to each sequence: 1) arrival time, 2) congestion time, and 3) type of traffic information. We used three levels of information: A is today's information, B is (in addition to A) information about driving speed on a chosen road section and C is (in addition to A and B) information about fastest route when all alternative routes are taken into consideration.

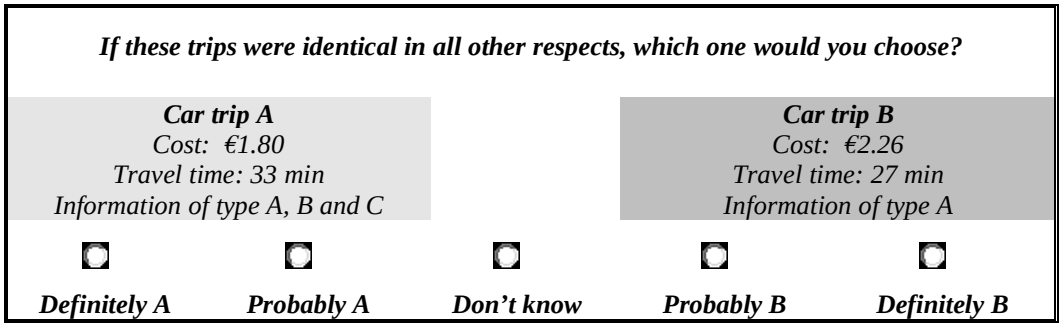


Figure 2. Example of a stated choice sequence. Traffic information study

In the public transport survey there were four choice sequences with six choice situations in each sequence. The respondent had previously been asked to describe a particular journey they've made by public transport. The cost of the journey was included in all the sequences. In three of the sequences the alternatives were bus and the mode which the respondents had used last time they travelled with public transport. In the last sequence the alternatives were car and bus.

The attributes included in the first three sequences were: Cost, time to the bus stop, headway, travel time, transfer, delays and travel time with or without a seat.

In the last sequence with the alternatives bus or car, the attributes for public transport was cost, travel time and headway. For the alternative "car" the attributes were the cost and travel time.

3. Why an Internet survey?

In Norway a large proportion of the population (58 percent in 2002) have access to Internet, and the Internet skills are generally good enough to answer an Internet-based Stated Preference study. In Great Britain the proportion in 2002 were only 38 percent (TNS 2002).

Internet allows complex questionnaires to be administered more quickly, flexibly and inexpensively than conventional survey methods. However, Internet is restricted to individuals with access to this medium. The two surveys mentioned in this article where carried out in the Oslo region. The access-rate (from home) in this region is 67 percent and 43 percent of the population use Internet every day (Vaage 2003). In addition all public libraries in Norway offer free use of Internet, so theoretically all Norwegians have access to the Internet.

Considering the high access rate, the relatively large proportion Internet users, and the fact that the access rate is growing, we find it very interesting to develop Internet-based surveys. However the access rate vary between population groups. Among the oldest people the access rate is poorest, and those who have access don't use Internet as much as younger people. In general, people with low income and/or only basic education don't use Internet as much as those with higher education and/or high income. To get a representative sample, we must bear in mind that Internet as the only source may give biased results. To avoid this problem it is possible to use other methods, like home interview or a paper copy returned by mail, as a supplement to Internet (Fearnley and Sælensminde 2001). In the two surveys described in the current paper a paper copy version was used as a supplement.

Traditionally, home interviews with an interviewer have been the way to collect data in this kind of surveys. Internet is not as expensive and has some additional advantages. Respondents who use Internet may answer the questionnaire whenever they like, day or night. They can use as long time as

they like and take a break if needed. This may cause better understanding of the questions, and improve data quality. Internet allows pop-ups and hyperlinks to give detailed instructions. However there is no one to ask if they need someone to explain something more properly and there seems to be a tendency to ignore what isn't understood immediately. To get high quality replies on Internet the design must be totally self-explanatory.

Data collected on Internet is ready to analyse immediately. The data don't need to be processed and in this way you can save time and money. On Internet you can follow the replies from minute to minute. This gives you the opportunity to look at the results and evaluate the balancing of the attribute levels in the *Stated Choice* **during** the answering period. If the balancing isn't good enough, you may change the design and let the rest of the respondents use the "new" questionnaire. In the pilot survey, this can be a useful way to find better balance *Stated Choice sequences* to use in the main survey.

Some respondents pay for their Internet access. This means that the respondents have to pay an amount of money to be able to answer the questionnaire. This cost may give an undesirable effect and influence the response rate. Still we know that many individuals access Internet free at charge at work, at school or at the public library.

As all computer based methods, surveys on Internet derive advantage from skip patterns, randomisation of questions and customized design. Customized design makes the attribute levels more familiar to respondents and provides higher variation in the observations.

4. Traffic information survey

Previous to the main traffic information study a pilot study was carried out in 2001 with testing of different methodological design for revealing travellers' valuation of traffic information. *Norsk Gallup Institutt AS* recruited these persons through a telephone survey. The interviews were conducted at the respondent's home, and the interviewer used a laptop PC. This enabled customized choice situations to the respondent's actual journeys during the interview. In our main study we needed at least 200 respondents. The cost of one successful home interview is around 1100 NOK. Using home interviews would then become very expensive.

In our survey, the target group were people who commute by car to work in Oslo and experience congestion problems on their route. We assumed that Internet access is quite high in this group, and decided to put our questionnaire out on a web site. It would only be accessible with a password provided by us. We also designed a paper copy version of the questionnaire in order to make it possible for those without Internet access to participate in this study.

One morning in September 2002 during rush hours nearly 1800 cards were distributed to car users at three locations in Oslo. Hence we recruited participants along the road during morning rush hours. The main traffic flow towards central Oslo in the morning follows the three main corridors: From the west, the northeast and from the south.

The cards explained a little about the survey and gave the web site address to the questionnaire and the individual password that would provide access. A maximum of two trials for logging in and participate were possible per password. There were no possibilities for one respondent to view the answers of other respondents. Those who wanted to receive a paper copy questionnaire could fill in name, address, trip distance, usual travel time and congestion time on the back page of the card, and return it to us free of charge. They would then receive the questionnaire a few days later. This was an

intentional barrier to using paper copies: We preferred a high percentage of participants using Internet. We also wrote on the card that was distributed, that one participant would win a bicycle, and that by answering on the Internet your chances of winning would be doubled. In addition, we reminded them that Internet is available in any public library.

Five different versions of the paper copy questionnaire were made. They differed with respect to travel time and distance. Based on a few data from the receiver we could then send the version that fitted best in each case.

It takes about 25-30 minutes to go through the questionnaire regardless of respondents used the paper copy questionnaire or answered on Internett. We got feedback from some respondents that the questionnaire was too long and difficult, especially the Stated Choice sequences.

The experience with Internet surveys at TOI was limited previous to the main study. Considering the comprehensiveness of our questionnaire there was a large degree of uncertainty related to response rate as well as to how representative the sample would be. In total, 1735 cards were handed out and 314 persons participated on Internet. Some respondents only completed the first choice sequence. About 10 percent of those who started on the questionnaire did not finish it. 278 completed the whole questionnaire. This gives a response rate on Internet of 17.5 percent.

It turned out that about 2/3 of the respondents, or almost 200 persons, responded the same day as they received the card (Figure 3.). About 40 responded the day after, and then there were only a few responses daily during the 12 remaining days. As can be seen in Figure 4, most of the answers came during the first hours after the distribution of the cards. It seems most people answered during job hours and very few from home.

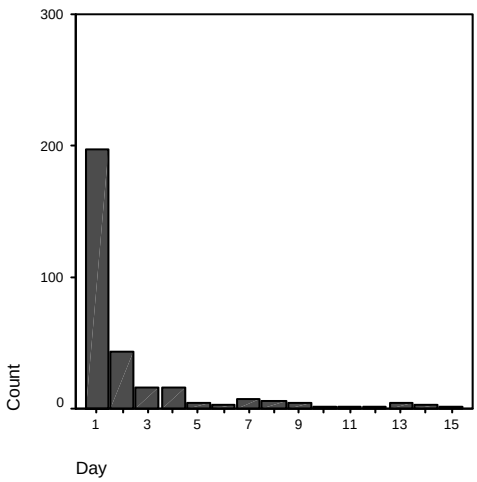


Figure 3. Responses per day on Internet

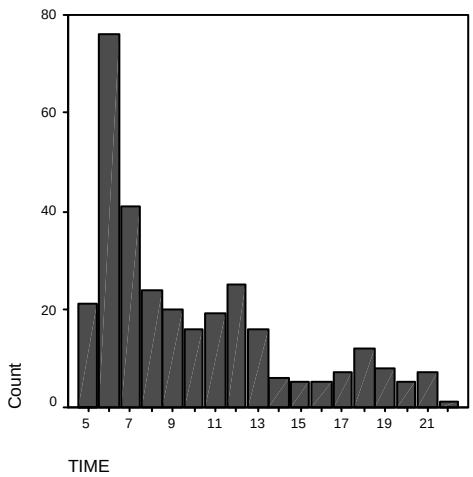


Figure 4. Responses on Internet during the day

Further, only 23 persons made use of the possibility to return the card and get an adjusted paper copy questionnaire by mail. Of these, only 7 returned a completed questionnaire to us. This data set was too small to be included in the analysis, and hence they have been ignored.

Because of the way the participants were recruited, we do not know the identity of those who received the card. Hence we had no possibility of sending reminders.

Because we chose to use the Internet, we were eager to see if that could lead to a biased sample. It turned out that ¾ of the respondents were men and that there were only a few respondents in the age groups 20-29 and 60 and above. Comparing these figures to the *Norwegian Travel Survey* (Denstadli and Hjorthol 2002), extracting data about gender and age for those who work in Oslo and use the car

as their main mode of transport to work, we could see that the characteristics of our sample fit well with the travel survey, both with respect to age and gender. Using Internet does not seem to give biases in these respects, given the fact that our target group was people who drive to work.

By using Contingent Valuation questing we calculated willingness to pay for reduced travel time, reduced arrival time variation (or travel time variation), reduced time spent in congestion and better information for car users driving to work.

Table 1. Willingness to pay for improvements of the trip to work. Traffic information survey

Factor	Willingness to pay
Reduced travel time	€5.17 / h
Reduced expected delay	€7.76 / h
Reduced congestion time*	€8.02 / h
Traffic information of type B	€0.41
Traffic information of type C	€0.54

**: 18 respondents who say they do not usually experience congestion are not included*

The valuation of €5.17/h for reduced travel time is in line with the recommended value of time for car trips to work based on the Norwegian value of time study (Killi, 1999). The willingness to pay of €7.76/h for reduction in expected delay implies that reduction in expected delay is valued fifty percent higher than travel time in our study. A higher willingness to pay for reduced travel time variation than for travel time in general has also been found in other studies (Small et al., 1999). Regarding valuation of reduced congestion, it is usually assumed that it is around twice the willingness to pay for reductions of not congested travel time. In Table 1, we can see that in our case it is significantly lower than twice as high. This, together with the fact that a large part of our respondents revealed no willingness to pay at all for reduced congestion, indicates that our sample has a lower than usual willingness to pay for this factor. A possible reason for this could be that the respondents were recruited directly from congested traffic and might on average represent a less congestion-averse fraction of travellers.

Of those who revealed no willingness to pay for improved information 35 percent, or 18 percent of the total sample, said they found the existing level of information satisfactory. This group represents about 18 percent of the respondents. If we also take into account the fact that what to expect regarding traffic conditions on the trip to work usually is well known to the traveller, we would not expect to find a high willingness to pay for improved information. Studies of other trip purposes would probably find higher values. There are not many other studies to which we can compare our results since this is a relatively new field. The values of €0.41 and €0.54 for information types B and C might be quite low, but on the other hand low values are what we should expect. Level A might be good enough for most travellers. Until better estimates are provided these are the values we recommend for use in cost benefit analysis of traffic information measures.

The intension was to calculate willingness to pay for the different factors by the Contingent Valuation method as well as by the Stated Choice method (SC). In our study the results of the former technique should be given more weight, as the Stated Choice sequences to some extend seem not to have worked out as intended. The reason for this is probably the complexity of this part of the questionnaire. The same design worked well in the pilot study, but then the respondent could ask the interviewer if something was unclear. In the main survey they were on their own, and there might have been a tendency to ignore what they could not immediately understand. Other reasons for the difficulties with calculations of willingness to pay could be that the factors travel time and cost might have dominated

the factors travel time variation, congestion and information type in the respondents' choices between trips A and B. Consequently, the chosen levels of the latter factors will not reflect actual preferences with respect to those factors.

From the SC sequence with travel time, cost and congestion it was possible to calculate valuation of reduced congestion. However, the value was lower than the value of travel time. From the SC sequence with travel time, cost and information type, the following values were found: Valuation of improving information from type A to B was about €0.12/h, and from type A to C about €0.33/h.

5. Public transport survey

The sample (5000) was chosen randomly from the population register. In that way everyone with address in the method area had a nonzero chance to be included in the survey. However those selected were aged 14 years or above. All had home address in the capital area (Oslo and Akershus). The data from the public register contained information about names, addresses, year of birth and gender. Respondent that were selected received a letter by post with an Internet address and a personal user name/password to log into the survey. Those who did not have access to the Internet were given the opportunity to fill in a paper copy version of the questionnaire. The paper copy was enclosed in the same envelope as the Internet address and user name/password. As well as the paper copy, a prepaid envelope was enclosed. In that way the respondent wasn't charged any money to reply. The letters were sent by non-priority mail, and therefore the sample received it over a period of tree to four days. The intention was to save money (postage) and to avoid collapse in our server.

After two weeks we sent a reminder by mail to those who hadn't answered. We repeated the Internet address and the personal user name/password. We wanted as many as possible to reply on the Internet, hence we only repeated the Web-address and didn't repeat the alternative way to do it (by pen-and-paper). The reminder influenced the response rate. The response rate declined after few days, after two weeks the reminder was sent by priority mail and the response rate rose again (Figure 5.1).

Many respondents preferred to answer the questionnaire on Internet in the afternoon (Figure 5.2), as opposed to the traffic information survey (Figure 4.).

The response rate was 30 percent, 14 percent used the paper copy and 16 percent used the Internet. The analysis is based on 1640 replies. In 1992 a similar survey was carried out in Oslo. Then they made home interviews, using portable personal computers. The response rate from that study was 28 percent (of those who were contacted) and the analysis was based on 644 replies.

The replies from 2002 were relatively evenly distributed between Oslo and Akershus counties and between the genders. The younger age groups used the Internet more often for their replies. Those aged 60 or above preferred to use the paper copy, but still there are some people over 60 year who replied on Internet. Men used Internet for their replies more often than women. Students/pupils and people who are working do reply more often on Internet than others. Recipients of benefits and retired people preferred the paper copy version. This shows how careful you must be with Internet surveys in order to avoid biased results.

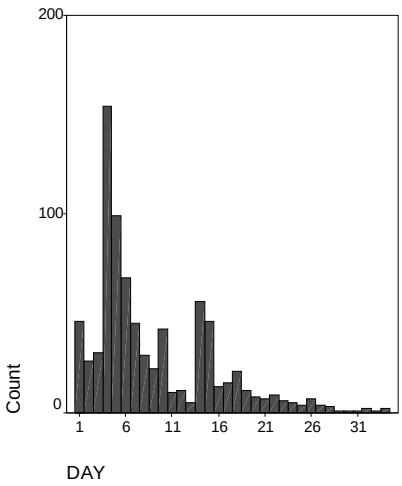


Figure 5. Responses per day on Internet

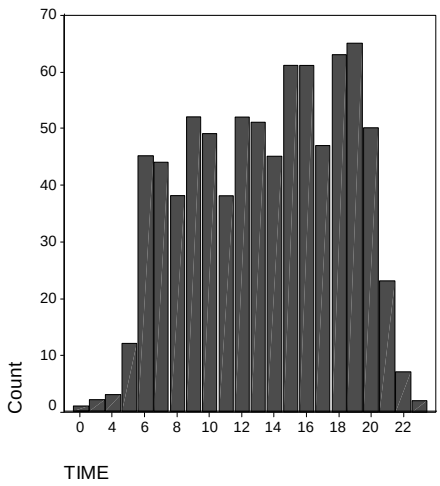


Figure 6. Responses on Internet during the day

In the pilot project the paper copy questionnaire was not as comprehensive as the questionnaire on Internet. The paper copy questionnaire turned out to be more easy to use, and we believe that it is the reason many preferred the paper copy instead of Internet in the pilot study. To encourage the respondents to use Internet, the paper copy was made longer and more complicated, and more similar to the questionnaire on Internet in the main study. At the same time we decided that only the Internet users would bee able to win the main prize (approximately €1300). All participants would be able to win one-month season ticket on public transport. The result was that a small majority used Internet in the main survey.

We were afraid that the questionnaire was to long and complex. So, in the pilot project we decided that some of the respondents on Internet were given a longer and more complicated questionnaire than others. On average it took 20 minutes to answer the long questionnaire on Internet. We couldn't see that the length and complexity affected the response rate, the time the respondents used to fill it out or the quality of the data, so we continued with the long and complex version.

The first time we ran an Internet-based survey we established a help desk. We learned that most of those who called had basic questions, almost none of the calls were concerning trouble with our server etc. We didn't continue the help-desk and for the next surveys we relied on e-mail. It turned out to be a good enough solution.

In the paper copy version there was a question about why they didn't use the Internet, and in the Internet questionnaire there were a question about how Internet worked. Half of those who used the paper copy said that they didn't use the Internet because they didn't have access at home or at work. 6 percent said they used the paper copy because they had problems connecting to the Internet. Some may have had technical problems logging on to Internet, and not answered the survey at all. We don't know anything about that group. Still the low number of questions received by phone and e-mail indicates that it can't bee a major problem. Those who used the Internet said that it worked well technically. However they found the questionnaire to long, complex and tiresome.

From the population register we know the name, gender, address and age of each respondent. This gives us a lot of information concerning those who answer and those who don't. However if the respondents failed to write down the password/user name in the paper copy we don't have this information. The paper copy included a question about the respondent's age. This was meant as a control variable to make sure that it was the right person who answered. It turned out to be quite a good match with age data from the public register and the questionnaire. Without a password and user

name you were not able to use Internet, hence all the answers on Internet has information concerning gender, age, name and address.

No upper age limit was set. However we do see that the quality of the answers is decreasing after a certain age. Individuals who are 70 year or older are excluded from the analyses since we don't find their answer of good enough quality.

On Internet the respondents was able to answer the questionnaire twice. Those who have logged into the questionnaire more than once had in most cases been interrupted the first time and returned to completed the second time. If they completed twice the first was chosen.

The similar survey from 1992 was carried out using recruiting by telephone followed by a home interview. The same attributes were analysed and the only effect, which we believe is from the change in method, is the reduced variance in the observation in the new survey i.e. improvement of data quality.

6. Experience and conclusions

The Internet was used as a tool for carrying out our studies. Using the Internet in this manner made it possible a) to produce customised questions for each respondent, in the sense that figures presented in some questions could depend on answers from previous questions, b) to enforce routing by not asking questions which the respondent was to skip, and; c) to make the collected data immediately available for analysis. These advantages could also have been obtained by carrying out home interviews, where the interviewer brought a portable PC. But that is a far more expensive method.

In the traffic information study it would have been more than twice as expensive to use home interviews. Each home interview costs approximately €140, including €35 for recruitment and €105 for the interview itself, while putting the questionnaire out on a web site will cost about €50 per respondent (even assuming that we have to over-recruit by 50% because of the higher dropout rates with an Internet as compared to a home interview survey)

In addition, we can use other ways of recruiting people when we decide to put the questionnaire out on a web site. That makes the survey even less expensive. How we choose to recruit people will depend on type of survey. For instance, in our public transport survey, described in chapter 5, we drew a random sample of the population in the survey and recruited people by surface mail. This was inexpensive and well fitted for this type of survey. In the traffic information survey, described in chapter 4, we recruited participants along the road during morning rush hours by delivering out cards. Since we wanted to reach car users who experienced congestion problems on their way to work, this was a suitable way to recruit people. In both these cases the chosen method for recruiting people is far less expensive than to recruit by telephone. Recruiting people for an Internet survey may therefore in some cases, be less expensive than to recur for home interviews.

The cost is important when choosing how people ought to be recruited but there are other aspects that are even more crucial. The main issue here is that how we recruit people also have to be considered in each case due to survey design, methods chosen and target group. For instance, in the traffic information survey described in this paper, our data set seems to imply that our sample of respondents has a lower willingness to pay for reduced congestion than the population average. This most certainly has to do with how people were recruited in this survey.

When you decide to put your questionnaire out on a website it is important to make the questions suitable for an Internet survey. In particular, we would emphasise that a complex questionnaire that seemingly works well for the purpose of home interviews might not work as successfully in Internet surveys. Too long and complicated questionnaires might lead to respondents giving up or not understanding the questions. The analysis of the Stated Choice sequences, in the traffic information survey (chapter 4), suggests that the questions were too complicated and that this have affected the results. In home interviews the respondent can ask the interviewer if something is unclear. During Internet surveys respondents are on their own and there may be a tendency to ignore what they cannot immediately understand. Therefore it is important, when designing an Internet survey, to keep in mind the facts that the questions should not be too complicated and the length of the survey should be limited. It is also important to emphasize the fact that questions should be self-explanatory.

On the other hand, compared to home interviews, Internet surveys can also have positive effects. The respondent can sit in peace and quiet, without being disturbed by and interviewer who may affect the answers given and cause bias. On Internet the respondents can use as long time as they want to answer the questions. In the public transport survey (chapter 5) we have experienced that the quality of the data has in fact increased when basing the survey on the Internet.

When considering using an Internet survey it is important to evaluate whether you will get a representative sample since everybody does not have Internet access. As we could se in chapter 3, Internet access from home increases with respect to income and education. Compared to other countries the Norwegian population is quite homogeneous due to the items just mentioned. With exception of the age group above seventy, the probability of having access to Internet at home was between 60 and 70 percent for all age groups above ten years (Vaage 2003). In addition, in Norway the Internet is available in any public library. Norway probably has a higher access to the Internet than most other countries, but Internet access in many countries is currently growing very fast.

Again, you have to consider whether to use an Internet-based survey in each case. In some studies you are only interested in a subgroup of the population. If that is the case, you have to consider how the Internet access is in the subgroup. In the traffic information study (chapter 4) we were only interested in people who commute by car in Oslo and experience congestion problems on their way to work. In this group we considered the access to Internet to be high because: The level of income and education in Oslo is quite high, and the target group is people who work. It is reasonable to believe that most of them have access to Internet either at home or at work. This is an example of a study where it seemed suitable to use an Internet survey. In fact it turned out that a large proportion of the respondents answered the questionnaire on the Internet during working hours.

In some other cases this might not be as evident as in the traffic information study. Then a recommended solution could be to use an Internet survey in combination with home interviews or a paper copy version of the questionnaire for those who do not have access to Internet. What will be considered best in each case will again depend on conditions like cost, target group and survey design. The advantage of using home interviews in combination with Internet survey is that in both cases it is possible to use customized design if that is desirable. In addition, the collected data would be ready for analysis immediately. If customized questions are not necessary in the survey or if it is important to keep the costs at a moderate level or to reach a large share of the population, using paper copy version in combination with Internet could be the recommended method.

A challenge when using Internet survey is to make it attractive for the respondents to participate by answering on the Internet. It is important that the survey has a subject that engages people or at least

introduce the questionnaire in such a way that people get interested. In general it is also an advantage if the questionnaire is easily accessible. The possibility to win a prize if participating also seems to affect the participation. If it is desirable that as many as possible answer by using Internet, and there is at the same time possible to answer in other ways, making it more likely to win by answering on the Internet may increase the percentage of respondents using Internet.

We have described experiences and tried to draw some conclusions from the two first Internet-based studies carried out at TOI. The results from these two studies have been promising. Still we have a lot to learn from this type of studies. During the spring 2003, new Internet-based studies will be carried out at TOI and hopefully the results from these studies will develop our knowledge further.

References

Bates J, Polak J, Jones P and Cook A (2001) The valuation of reliability for personal travel. *Transportation Research* Part E 37 (2001) 191-229.

Brand D (1998) Applying Benefit/Cost Analysis to Identify and Measure the Benefits of Intelligent Transportation Systems. *Transportation Research Record* 1651, Paper No. 98-0938.

Denstadli J M and Hjorthol R (2002) *2001 Norwegian Travel Survey – key results.* Oslo, Institute of Transport Economics. TOI Report 588/2002. ISBN 82-480-0277-2.

Fearnley N and Sælensminde K (2001) *Testing Stated Prefernce techniques and stated choice design: A methodological study of public transport passengers' valuation of quality.* Oslo, Institute of Transport Economics. TOI Report 544/2002. ISBN 82-480-0229-2.

Killi M (1999) *Recommended values of time for passenger transport.* Oslo, Institute of Transport Economics. TOI Report 459/1999. ISBN 82-480-0123-7.

Killi M and Samstad H (2002) *Travellers' valuation of traffic information based on work journeys.* Oslo, Institute of Transport Economics. TOI Report 620/2002. ISBN 82-480-0314-0.

Nossum Å (2003) *Public transport in the Osloregionen. Travellers vaulation of time*. Oslo, Institute of Transport Economics. TOI Report 633/2003.

Small K A, Noland R, Chu X and Lewis D (1999) *Valuation of Travel-Time Savings and Predictability in Congested Conditions for Highway User-Cost Estimation*. National Cooperative Highway Research Program, Report 431. National Academy Press, Washington, D.C.

Sælensminde K (2000) *Valuation of nonmarket goods for use in cost-benefit analyses: Methodological issues*. Oslo, Institute of Transport Economics. TOI Report 491/2000. ISBN 82-480-0166-0.

TNS Interactive – Global eCommerce Report – June 2002. Taylor Nelsen Sofres PLC 2002.

Vaage O F (2003) *Norwegian media barometer 2002*. Oslo, Statistics Norway, Statistical Analyses 57

About the Authors

Marit Killi and Åse Nossum are both Research Economists at the Department of Transport Economics at the Institute of Transport Economics (TOI). TOI is a national institution for transport research and development in Oslo, Norway. They can be contacted via Marit at Institute of Transport Economics, PO Box 6110 Etterstad, N-0602 Oslo, Norway; tel. (+47) 22 57 38 00; fax +(47) 22 57 02 90; e-mail mki@toi.no.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

The Use of SMS as a Research Tool

Mike Cooke, Anders Nielsen, Colin Strong

Abstract

In this paper we examine whether or not SMS is likely to be a "disruptive technology" in the field of marketing research. After an extensive programme of research we conclude it is likely to remain, at present, an auxiliary research tool that will facilitate our speedy and emotionally satisfying interaction with a significant minority of the population.

1. Introduction

A "disruptive innovation" is one that creates an entirely new market though the introduction of a new kind of product or service. According to Christensen, a fatal threat to established companies' market share can begin as a low-quality, low-margin product that your customers don't want and can't use - yet. A disruptive technology is a new product or service that is not necessarily as good as your current product line. Typically, it's technologically simple. Often, it's more convenient to use. It's something that's cheaper, simpler, and often smaller.

Short message service (SMS) text messaging is certainly a simple advertising medium. It is limited to 160 text characters, has no sound or high quality visuals. But ignore these disruptive technologies, and they just may grow in capability to meet mainstream needs. SMS has done just that. It has revolutionised inter-personal communications. From being a way to exchange information between engineers it has revolutionised person-to-person communications and entertainment. Can it do the same for advertising and marketing research?

The explosion in personal text message usage which has taken place in recent years has led to ripples of excitement amongst many researchers because of the new research opportunities it represents. The growth in mobile telephone ownership has been dramatic with overall penetration at almost 77%, reaching near saturation in some demographic segments. For example 80% of 18-24 year olds use mobile phones and over 90% of this group make use of text messaging.

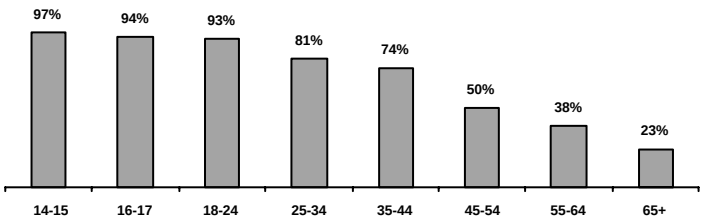


Figure 1. Text messaging usage by age group

Base=All UK mobile users (N=1,000)
Source = NOP Mobile User Study

What started as personal Consumer-to-Consumer pursuit is increasingly moving towards a commercial, Business to Consumer activity. One rapidly emerging area is text-based advertising, whereby third parties use the 160 characters available to communicate with their target market. This completely new advertising medium is a very exciting one in which a number of companies are competitively selling advertising inventory and FMCG brands are keen to explore the possibilities of the medium and understand how they can best use it to promote their brands effectively.

To facilitate their inventory sales these new sales companies have had to build large databases of mobile phone owning individuals. These databases may contain detailed behavioural and demographic profiling information to allow precise targeting of their client's advertising offerings. This in turn has created the possibility of a new research medium that allows surveys to be delivered by telephone to precisely targeted samples of the mobile population, with their own personal return path. The potential cost and time advantages are obvious. A number of the ASP companies are now actively marketing the research opportunities that these 'panels' offer for conducting market research.

The emergence of this new channel poses both opportunities and challenges for researchers in terms whether the medium creates new opportunities as a channel for research.

2. Research programme

Willingness To Participate

Our research programme examined the degree to which mobile phone users, identified from our online research panel, would be willing to cooperate in market research studies that were conducted by SMS. Initially we added a number of questions about SMS to one of our online panel calibration studies and were not encouraged by the results.

- From a sample of 4,661 members of the NOP Online panel we identified 3,474 who used the SMS capabilities on their mobile telephones. Only 19% of these SMS users granted us permission to send them surveys using SMS. Given that these were NOP online panellists who had agreed to take online surveys, we were surprised that the cooperation level was so low.
- Consequently, we undertook an online qualitative study, with 500 of the panellists who had refused to participate, to ascertain their reasons. A major finding was that these people were still willing online panellists as 382 accessed the survey and 341 completed it. This was a 76% response rate and an 89% completion rate. It was not participation in surveys that was the issue for these people but rather that they were unwilling to respond via their cell phone using SMS.

The major reasons for their unwillingness to participate were:

- The personal relationship that they have with their mobile telephone
- The intrusive nature of SMS
- The perceived cost of receiving our surveys & of sending their replies
- The volume of SMS Spam
- The limited in-box storage capacity on mobile phones
- The belief that online surveys are more convenient

The following give a flavour of the responses:

> *"My mobile phone is for friends and family to contact me at any time of the day and not a marketing tool. I find it a real nuisance companies contacting me via the mobile for promotions and marketing updates. For this purpose alone I never sign my number to anything, simply because I know that old databases can be bought and eventually the promotional materials will start ringing there way to me. Sorry - I just find it a real pain and an anti-climax...when I hear an*

alert sound I immediately assume it's one of my nearest and dearest and prepare to read in excitement - only to find it's ANOTHER promotional offer!"

"I consider my personal phone to be for my personal use only - especially with a limited size inbox I don't want the hassle of clearing unnecessary messages that I don't want. I want to keep at least one comms medium private and free from Spam"

"I prefer to answer when I have the time rather than be interrupted. Email gives me control over when and where."

"I find SMS intrusive, don't like the screen size and find it unnecessarily fiddly to input text replies."

"Because the questions are likely to arrive at inconvenient times such as when I'm at work, and it will cost me 10p a time to reply to them! If I answer the questions over the internet, it costs me 1p a minute at the most."

"… because I don't want to be disturbed by my phone beeping unless someone I actually know has something important to say to me, because I don't want to feel I have to sit and type out a text message to you when I'm in the middle of doing other stuff. My phone is there as a useful tool not as another way to bug me."

However it was not all bad news as we discovered that twice as many SMS users, from our online panel (38%), were willing to receive SMS messages from NOP notifying them that an online survey was ready to complete. While this is much more limited role for SMS in survey research it is be a valuable addition to our armoury. The personalised nature of the mobile telephone is such that a call to action is unlikely to be ignored and an SMS in conjunction with an email could be a powerful incentive to complete our online studies.

We were also encouraged to find that the profile of panel members willing to participate in SMS surveys and those who would accept SMS alerts did not differ from the SMS using universe.

It was clear from the qualitative results that we had the opportunity to address some of the major obstacles to participation at the opt-in solicitation stage. We decided to select another random 5,500 members of our online panel who used SMS and request their permission to send surveys by SMS but this time we would emphasis that NOP would cover all costs involved in sending and receiving messages, that we would reward participation, that the surveys would be short, and that they would be only sent occasionally.

The effect of these changes was a very significant 42% increase in the proportion willing to participate in our SMS surveys. It rose from 19% at stage one to 27% at stage two, once we had addressed our panellists concerns. And again we found that the profile of panel members willing to participate in SMS surveys did not differ from the SMS using universe.

The lower than expected levels of cooperation, likely to be found in traditional research panels, will clearly be an issue with SMS, as an interviewing mode, unless we identify those who are willing to participate in advance and foster their participation. Permission will be as important in research as it is in advertising.

It is also worth putting the response rates into the wider context of falling cooperation levels generally. We did this by examining the willingness of our online panel to take surveys by two traditional methodologies namely telephone and mail. What we found was interesting. In the UK 76% of the panellists were willing to receive surveys by mail but only 16% were willing to take them by telephone.

These findings are in line with the latest "2001 Respondent Cooperation & Industry Image Study" produced by The Council for Marketing & Opinion Research, which argued that consumers prefer Internet and mail surveys to the more intrusive telephone method. In the last study

CMOR asked 2,000 people, b*y telephone*, what their first-choice survey method was and 40% selected "Mail", with "Telephone" chosen by only 26%.

Further, when they asked 1,700 people, *via the Internet*, what their first-choice survey method was, a 79% majority chose the "Internet" while only 3% chose telephone. Most interestingly "Snail Mail" was preferred by four times as many as preferred the telephone.

So SMS & traditional telephone interviewing have similar cooperation levels among our online panellists but obviously once someone has agreed to participate the telephone is a more flexible interviewing mode. Unlike SMS it is not limited to 160 characters per message. Also the telephone is a proven methodology for conducting long as well as short interviews. Our believe is that the very strengths of SMS, its personal nature, precise targeting and portability, which make it such a powerful advertising medium, may actually work against it as a research medium. Our panellists see it as an intrusive survey methodology over which they do not have the control that they can exercise over online and mail surveys.

However, even if we argue that SMS is unlikely to be a uniquely "disruptive" research medium, it is possible that it can play a role in a multi-modal research design. In a world in which we want to encourage respondents to complete surveys by whatever mode is most convenient to them, it might be that there are certain research situations where we find that SMS has a role to play. For example Widman and Vogelius in a recent ESOMAR paper argued that it was a technique capable of measuring daily newspaper readership. They did this successfully by sending one question a day to their pre-recruited panel for two weeks.

SMS as a Research Medium: The Test

In stage three of our research programme we explored the types of research that might be conducted using SMS. This study was an experimental design in which we tested the effects on SMS survey response and completion rates of asking different types of questions, using different interview lengths, and for different durations of time.

There were four modules in our test. In the first module respondents were asked to complete 3 questions for 5 days. The questions were in the form of a "Day After Recall Diary" in which respondents were asked to record how many calls they made and received on their mobile telephones and how many SMS messages they sent "yesterday". The second module consisted of 3 quite "Personal" questions, on one day, including questions about religion, work absenteeism and voting; the third module was of four "Newsworthy Opinion" questions, on one day, about terrorism, fox hunting, football and movies and the fourth module was of 5 "Standard Survey Research" questions based around use of the Internet and mobile telephony

The schema below summarises the research design.

Table 1.

Module	Type of Questions	Time Period (Days)	Number of Questions
1	Diary	5	3
2	Personal	1	3

Module	Type of Questions	Time Period (Days)	Number of Questions
3	Topical, News	1	4
4	Standard MR Survey	1	5

1,162 panellists, who had given their permission to be interviewed by SMS, were randomly allocated to these four test studies. They were pre-warned to expect the study and incentivised via our online panel points system. We made it clear that they would receive enough points to cover all their telephony costs as well as well as additional points to reward their participation. Panelists are supplied 'I-points' as an incentive for their participation in undertaking on-line surveys which can be redeemed for a variety of goods and services. The number of I-points given for this was equivalent to undertaking a normal on-line survey.

SMS as a Research Medium: Response Times

The research benefit that was immediately observed was the speed of response. We received over a third of the responses in less than 1 minute and over a half within a minute. 82% were received within an hour.

The table below shows the elapsed minutes from SMS dispatch to response:

Table 2

Time Elapsed in Minutes	Cumulative % Of respondents
Under 1 minute	36
1 Minute	58
2 – 5	63
6 – 10	69
11 – 29	76
30 – 59	82
60 – 119	89
120 – 240	94
4 hours plus	100

SMS as a Research Medium: Response Rates

Introduction

The overall response rate was 63% across the four tests. We were surprised and pleased at the high levels of participation across all four tests.

Table 3.

Module	Type of Questions	Days	Number of Questions	Response Rates
1	Diary	5	3	69%
2	Personal	1	3	61%
3	Topical, News	1	4	58%

Module	Type of Questions	Days	Number of Questions	Response Rates
4	Survey	1	5	64%
Overall				63%

Consequently, given these high levels of participation, we found that the demographic profile of the reporting sample closely matched that of the selected panellists. The following illustrates how close that match was in terms of gender and age:

Table 4.

	Issued Sample %	Reporting sample %
Male	50	49
Female	50	51
Age		
16 – 24	19	22
25 – 34	40	41
35 – 44	20	19
45 +	21	18

Response Rates and Survey Duration

Counter-intuitively the highest response rate, at 69%, was achieved in module 1, in which our potential respondents were asked to *"answer 3 questions via SMS texting every day for the next 5 days?"*

As the respondents did not know the nature of the questions at this stage, only the number of questions and survey time period, we have to conclude that response rates were not adversely affected by increasing the number of days over which the survey was conducted. This appears to point to the fact that SMS is a suitable research tool for conducting ongoing panel research where the respondent is required to record some aspect of their behaviour. This supports the findings of Widman and Vogelius who argue that it was a technique capable of measuring daily newspaper readership. Indeed it maybe more flexible than they imagined. In their test they sent one question a day to their pre-recruited panel for two weeks while we successfully sent three questions although for only one week.

Equally impressive is the fact that the number of responses achieved each day varied very little. There were 137 on day one, falling to 129 on day four before rising to 133 on day five. We also found a high degree of continual reporters, that is people who took the study on each of the five survey days. Three-quarters (73%) of those who agreed to take the survey on day one were still active on day 5.

However the duration of the study may have had a somewhat detrimental effect on the daily completion rates. Whilst we did not experience any variation in the daily completion rates, over the course of the study, the achieved rates were consistently lower than for the surveys completed on one day. In the 5-day study the percentage of respondents who took the study, and answered all 3 questions was consistently in the range 81% - 84% while in the one-day studies the range was 92% - 99%.

All these findings indicate that there was little sign of survey fatigue or "wear out", other than on daily completion rates, and again points to the suitability of SMS for panel based research.

In retrospect we wonder if offering to involve respondents over an extended period of time, plays to the "emotional" strengths of SMS. It is clear that the physical benefits of SMS as a research tool are easy to demonstrate. It does allow us to conduct surveys in an efficient and timely manner but it may be less apparent that this media facilitates extended "conversations" with respondents. SMS is a highly personal media whose emotional benefits are it's interactivity and the facility it offers to be always in touch. These are benefits that accrue to researchers using this methodology in panel studies.

Response & Completion Rates and The Number of Questions Asked

One of the key issues we wished to test was the degree to which respondents were willing to take more than one question. The SMS advertising community could not assist here. The general view was that SMS would work for one question, a la the advertising response model or the "Big Brother" type polling function, one sees on interactive television, but there was no advice on whether respondents would be prepared to take more than one question via SMS.

Consequently, we decided to trial the effect, on response rates, of informing people that we wanted them to take 3, 4, and 5 questions. Prior to the test we thought that 5 questions would be the maximum that people would be prepared to agree to take, let alone successfully complete. We expected to see a decline in response rates as we offered respondents more questions to complete. The results did not support this view. There was no evidence that response fell as the number of questions increased. Indeed at 64% the 5-question option achieved the highest response rate of the 3 one-day versions.

However the acid test is in the completion rates. How many of those respondents who take each of the studies actually complete all of it? Our results are instructive on this point. We found that the 3-question variant produced a completion rate of 99% indicating that once recruited it is feasible to ask at least 3 questions. We did experience a drop in completion rates when we increased the length to 4 questions. It fell to 93%. Interestingly the addition of the fifth question had little effect and a rate of 92% was recorded. These results have encouraged us to trial even longer questionnaires in the future to establish where the drop of in completion rates occurs.

Response Rates and The Type of Questions

We were interested to see if SMS was a suitable media for answering a range of different types of survey questions. In particular we debated whether or not we could ask quite personal questions or whether people would refuse to answer them. There were two schools of thought; the first that SMS was ideal for asking personal questions, as it is a uniquely personal interviewing tool and greatly benefits from using a self completion methodology, and the opposing view that thought it would be too intrusive to send such sensitive questions, over the telephone, without an interviewer to explain the purpose of the study and answer any questions that the respondent might have.

The 99% completion rate indicated that respondents were prepared to answer our questions about religion, absenteeism from work, and political affiliation. Indeed only 1 respondent took the pre-coded option and refused to answer the religion question and only 2 refused the absenteeism and voting questions.

Although the study was not designed to be representative of any population other than that of "our online panellists with mobile who had agreed to take surveys by SMS", the answers did appear to have face validity. We found that only a quarter claimed to be "religious", that three-quarters admitted to

having "taken a sickie" from work and that two fifths, recalled voting for the Labour Party at the last election. This last question also indicates the utility of SMS, as a data collection tool, as it showed that respondents were quite prepared to chose the traditionally less socially accepted answer of "Did not Vote" (25%) rather than opt for the less emotive "Rather not say" option.

Likewise, the "Topical News" questions also had face validity, with two thirds of the sample favouring the banning of fox hunting, almost half opposing a war with Iraq, while three-quarters were looking forward to the latest Harry Potter movie and less than 10% thought Liverpool would continue to go undefeated through the football season. (They lost to Middlesbrough within a fortnight of the study!)

Respondents' Satisfaction with SMS Surveys

One of the great functional benefits that comes with using SMS as a survey tool is that the user base is TXT literate. Users are used to typing their text messages in a concise way, making maximum use of the 160 characters available to them. Consequently it is an ideal vehicle for collecting open-ended responses to questions. We were never in any doubt that the use of open-ended questions would be successful but our most optimistic views were exceeded when only 3 of the 110 respondents who had the opportunity to reply to an open ended question failed to do so.

The question we asked them was *"What do you think about being interviewed via SMS like this?"* The overwhelming reaction was positive. Only 6 respondents gave a negative answer. Half of these simply stated a preference for answering via the web and they other 3 thought that SMS took too long. The other 94% of respondents were positive about SMS as a survey vehicle. The following quotes illustrate their views:

> *"I think it is an easier & better way to conduct surveys because it is faster, easier to complete & it doesn't mean going online when you aren't near the computer."*
>
> *"I thought this was a great way to be interviewed, quick and easy"*
>
> *"I think it's great, very easy and simple. I'd certainly do it again and again."*

However despite our assurances that NOP would cover all costs involved in sending and receiving messages the perceived cost of SMS was still an issue for 17% of our sample. The mechanic of replying to questions by sending TXT messages clearly induces a latent fear of the cost. This is probably an unfortunate side effect of some of scurrilous marketing practices that have occurred in this arena. Typical of these fears were responses such as:

> *"Not sure yet, it depends how much it costs me. Easy but prefer net as cheaper"*
>
> *"I would prefer not to have to send 5 messages as the is a cost for sending each one"*
>
> *"OK but it can cost a lot of money if u pay 4 txt as u go rather than having free minutes/txts on a contract"*
>
> *"It's ok for me as I get 500 free texts a month. But it wouldn't be so good if I had to pay for each reply"*

One respondent succinctly summarised the sample's views:

> *"Easy to do, at my convenience. More likely to receive notification than by email. Costs me money though and typing replies can be tedious "*

3. SMS: Research Implications

This research program suggests that SMS as a survey method appeals to a minority of our online panellists. The appeal is limited by fears about our research intruding on the panellists' private space and over the costs. However we were able to get much high participation rates once we addressed these fears head on, at the recruitment stage. Once they were recruited two thirds of our willing respondents participated in the experimental quantitative research test.

Not only did we achieve high response rates we did so in rapid time. We thought it was remarkable that a third of responses were received within a minute. This is a major benefit of this survey methodology.

We established that respondents would answer more questions than had been previously thought and we are able to support previous findings that SMS is a suitable research mode for conducting panel research. We speculate that by offering to involve respondents over an extended period of time, the researcher can play to the "emotional" strengths of SMS namely it's interactivity and the facility it offers to be always in touch. By offering the respondent an extended interview period we may be tapping into the emotional as well as the more physical and functional aspects of SMS.

4. Overall Conclusions

So is SMS likely to be a "disruptive technology"?

In the research field it is unlikely. We believe it will be a useful addition to our research armoury and facilitate our Martini inspired offering of providing NOP's panel with the opportunity to reply to our studies "anywhere, anytime, and any place" that suits the individual respondent. It will be analogous to the concept of "media neutral" planning in advertising in which our research communications are driven by our respondents' channel of preference, whether that be telephone, IVR, web, interactive TV or SMS. The challenge for the research community to understand how these research channels fit together and to comprehend the data that flows from this type of multi-modal interviewing.

References

Advani, Rajiv and Khaled Choudhury (2001). Making the Most of B2C Wireless. Business Strategy Review 12, 2 (Summer), pp 39-49.

Barwise Patrick (London Business School) **& Colin Strong** (NOP). Journal of Interactive Marketing Winter 2001 Permission –Based Mobile Advertising.

Barwise, Patrick (2001). TV, PC, or Mobile? Future Media for Consumer e-Commerce. Business Strategy Review 12, 1 (Spring), pp 35-42.

Christenson, Clayton & Overdorf Michael (2001) Meeting the Challenge of Disruptive Change. Harvard Business Review on Innovation

Christenson, Clayton (1997) The Innovator's Dilemma When New Technologies Cause Great Firms to Fail. Harvard Business School

Cooke, Mike.(2002) Multi-Modal Research in a Connected Society. Marketing Week. 6 June 2002

Council for Marketing & Opinion Research (2001). 2001 Respondent Cooperation & Industry Image Survey.

DMA (2000). 1999 Economic Impact: US Direct Marketing Today. Direct Marketing Association.

Enpocket. (2002) Consumer preferences for SMS marketing in the UK

Godin, Seth (1999). Permission Marketing. New York: Simon & Schuster.

Harvey, Fiona (2001). From Ibiza to 1 billion. FT Creative Business (7 August), pp 2-3. London Financial Times.

Strong, Colin (NOP) & Old, Joanne (Define) (2001) From intimate conversation to commercial communication: What are the opportunities and challenges for researchers given the changing role of text messaging? Research Magazine February 2001

Widman, Leif and Vogelius, Lena (2002). Daily Reach using SMS. Measuring and reporting reach in real time. ESOMAR WAM 12 June.

About the Authors

Mike Cooke, Anders Nielsen and Colin Strong all work at NOP and can be contacted via Mike at m.cooke@nopworld.com.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

SMS Polling. A Methodological Review

Joel Down & Simon Duke

Abstract

Two SMS (short message service/text message) surveys were conducted by MORI for BBC Watchdog in late 2002 and early 2003. This paper explores the benefits and constraints of SMS surveys including sampling, questionnaire design, response rates and data analysis. It shows that given the broad socio-economic reach of SMS messaging, SMS polling has the potential to complement other non face-to-face methods including internet surveys and CATI, especially when more advanced forms of mobile messaging are introduced. However, it cannot be used for representative general public polling until SMS usage is more common amongst older age groups.

Keywords

SMS; text message; opinion poll

1. Introduction

In July 2002 MORI was commissioned to conduct polling by SMS (text message) for BBC Watchdog. The original brief arose from a conversation with Doug Carnegie, Editor of Watchdog. He wanted to increase the opportunity for his audience to interact with the programme but was aware of the drawbacks of phone-in 'voodoo' polls and wanted to move beyond this biased form of research. Therefore he suggested SMS polling with subsequent systematic weighting of the findings to improve the accuracy of the data.

MORI therefore devised an alternate methodology, which involved:

- Recruiting a panel of SMS users
- Profiling the panel by a range of demographic variables
- Sending questions to panel members
- Weighting the findings

The research raised a number of methodological issues, in particular:

- How representative is a sample of SMS users vs. the GB population?
- What are the strengths and weaknesses of SMS as a research methodology?

2. The SMS Population

Before conducting the SMS polling, we needed to understand the profile of mobile phone and SMS users versus a profile of the population as a whole.

An analysis of general public data collected on the MORI Omnibus data in June and July 2002 showed that there were notable similarities in social class, voting intention, economic optimism and activism etc.

However, there was a dearth of SMS users aged 55+; although our Omnibus data showed that age was not necessarily a defining factor in responses to many questions, the absence of older people has a considerable impact upon weighting efficiencies. Evidence from lifestyle questions, where SMS users were more likely to hold liberal values, indicated that attitudinal weighting of some kind may be required.

With these issues in mind, we proceeded with an SMS pilot to test the concept.

3. SMS Polling Pilot

In August 2002, a question was placed on the MORI face to face Omnibus inviting respondents to take part in a text message poll for BBC Watchdog. 289 respondents agreed to take part. On Friday 30 August, all were sent two messages:

Message 1: *Last weekend you agreed to be part of an experimental poll run by MORI in association with BBC Watchdog. Later today you will receive a poll question.*

Message 2*: BBC Watchdog:How would u vote in a referendum on whether Britain should be part of the Single European Currency?Pls text A if in favour; B against; C don't know*

By the morning of Monday 2 September, 170 had replied.

- We had about 15 duplicates, which reduced the overall response level to 155, a response rate of 54%
- 140 of these responses were sent within three hours of the message being sent out at 6pm on Friday evening - suggesting that SMS surveys could generate a quick response.

Looking at the profile of the respondents, the age skew was evident. Only one of the respondents was over 65, thus with age weights their views were upweighted to represent 19% of the population. Without age weights, the findings were much closer to face to face MORI polling data.

Table 1.

How would you vote in a referendum on whether Britain should be part of the Single European Currency?	**MORI Face to Face Data**[†]	**SMS Data – Including Age Weights**[§]	**SMS Data – Excluding Age Weights**
In favour	31%	42%	31%
Against	55%	47%	52%
Don't know	16%	11%	17%
Base	(2,000)	(170)	(170)

[†] Weighted to be representative of GB adult population, July 2002. MFS Omnibus.

[§] Other weights were age within gender, social class of chief income earner, working status within gender and number of vehicles in household

4. Recruitment of Participants

Having reviewed the pilot and concluded the outcomes were positive, we agreed to go ahead and recruit a panel for BBC Watchdog.

One of the biggest challenges for SMS polling is generating a sample:

- Unlike RDD dialling, there is no single source of a mobile phone sample
- We cannot easily interrogate participants to collect demographic data, as we can only ask one question per message
- Therefore this is an opt-in survey: we need to collect demographic data through non-SMS channels before sending out the questions, plus the profile of those choosing to take part may be different to samples generated through other research methods
- Also, unlike other survey methodologies (perhaps with the exception of internet surveys) there is a cost to the SMS user when taking part – the cost of a text message

For reasons of time, cost and the desire for interactivity with Watchdog viewers, we decided to first recruit panel members through the programme. **This will impact upon the applicability of the findings to the general public**[1] but the overall methodological learning points still apply.

Initially a call to action was broadcast on Watchdog. This explained the purpose of the panel and invited people to send their name and address by text message to the BBC, or register on the Watchdog web site. Applicants completed an on-line or postal questionnaire that generated the demographic profile for each respondent later used to weight the data.

A total of c.2,600 panel members were recruited through this channel. After the first 1,500 responses had been received, it was clear that the over 55s would be under-represented. Although the 55+ group makes up 32% of the population, they made up just 6% of the Watchdog panel.

Furthermore, those aged 65+ are 19% of the population, but only 1% of the panel. Therefore we boosted the panel with older SMS users drawn from previous MORI surveys. This generated an additional 276 SMS users aged 65+.

The final profile of the 2,750 respondents in the SMS panel is attached as Appendix 1. In summary, sample cell sizes were sufficiently large on all standard weighting variables (social class, gender, working status, number of vehicles in household) except age.

5. Question Design

Whilst the panel was recruited, the first questions were drafted. The constraints on question design proved to be:

- The length of the question
- The types of question that could be asked
- The scales and precodes that could be used

[1] It is difficult to estimate how representative Watchdog panel members are versus the population as a whole. However, we have collected information on a series of demographic variables which allows us to profile this sample; and the questions we have analysed in this review do not relate to consumer issues *per se*.

Length

Currently, standard SMS messages are limited to 160 characters. This limits the length of questions that can be asked. As a result, for some questions we sent a preamble before sending the question itself.

Format

For reasons of data capture and analysis, we only used precoded questions. There is potential to use open-ended responses (whose input is not too difficult for those familiar with SMS, especially using predictive text messaging), but we have not trialed this with the Watchdog panel.

All of the questions asked are listed as Appendix 2. We used two types of question: a satisfaction scale and a single code question.

An additional challenge was to design questions that allowed SMS users to respond in a consistent format, to allow swift data processing. With the introduction of predictive text messaging, it is easier for many users to input a letter than a number; also, using numbers to reply to every type of question could affect some responses. Therefore all questions were written to generate a text response (A, B etc.)

Scales and Precodes

A **satisfaction** scale (for banking services) caused difficulties as we could not show the full range of the scale:

> *Pls vote on scale A to E where A=very satisfied->E=very dissatisfied, F=don't know*

Potentially it was unclear to respondents where the mid-point of the scale lay, thus was a response of C a neutral opinion, or positive or negative? This led to inconculsive survey findings.

Using **single choice precodes** was more effective, since there was no confusion over scales:

> *Do you think it's safe or unsafe for parents to choose the MMR vaccine for their children? Reply A=SAFE, B=UNSAFE, C= DON'T KNOW*

When composing the questions it was decided to offer a don't know option, which makes comparison with telephone and face-to-face methodologies a little more difficult when this is not an option presented to users. Whether the don't know code should be shown on an SMS survey should be treated on a case-by-case basis.

6. Fieldwork

One hypothesis tested in this project was that SMS polling would allow rapid response to survey questions, since most respondents are near their mobile for much of the day and SMS provides more intimate access to panel members.

With the first question sent at 1300 on Friday, more than three quarters of responses had been sent in the first two hours, and more than half within the first hour. A small secondary peak arose late in the following morning.

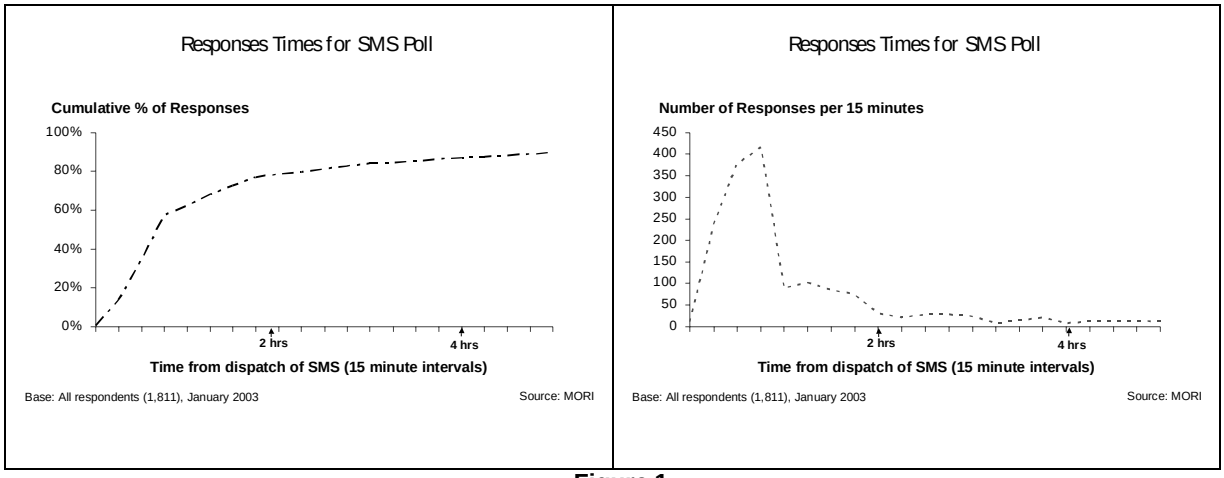


Figure 1.

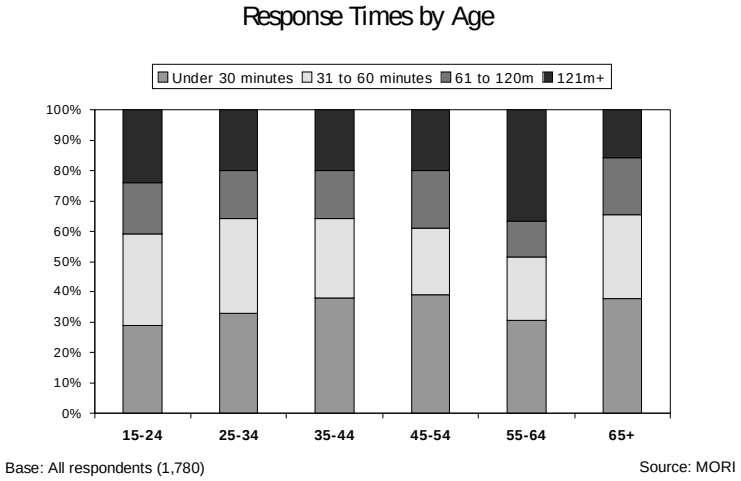


Figure 2.

Given their greater familiarity with mobile phones, it might be expected that younger people would be the quickest to respond. This was not proven in the survey, where the speed of response was similar across all age groups.

Difficulties arose with the pilot poll when the gateway was not able to handle the volume of SMS received, but a different SMS provider was used for the main survey and there were no capacity problems.

When sending more than one message for a survey, the order of receipt is clearly important; sufficient time should be left between the dispatch of each message to ensure they are received in the correct order. For example, if a user turns their phone on and receives a series of text messages, with at least some operators and/or models of phone the most recent message is received first – which would reverse the order in which multiple survey messages are dispatched.

7. Data Processing

The data, were supplied by the SMS provider in an Excel format with information including the order of response, time of response, respondent's phone number and the full text of their message.

Since respondents are used to sending SMS messages, rather than single letter responses, many responses required cleaning. A handful of duplicates had to be removed, and text responses required

interpretation and cleaning. For a survey of this scale this added around an hour to the data processing stage; for larger surveys, more time would be required.

Once the data had been cleaned, text codes were converted to numeric values and the data were matched with the panel profile data using the telephone number as the unique identifier.

8. Data Analysis

Data were weighted using the same variables as the MORI face-to-face Omnibus:

- Age within gender
- Social class of Chief Income Earner
- Working status within gender
- Number of vehicles in household

The shortfall in those aged 55+, especially 65+, significantly reduced the weighting efficiencies of the survey.

9. Validation of Findings

Appendix 3 compares the SMS survey findings with existing MORI Benchmark data.

Responses to the December 2002 survey on **satisfaction with banking services** seemed to show a higher level of dissatisfaction than amongst the public as a whole as measured on the MORI Financial Services Omnibus (9% dissatisfaction on SMS survey versus 2% dissatisfaction on MORI benchmark). We cannot be sure of the reasons for this, but they are likely to be a combination of:

- The profile of Watchdog viewers who agreed to take part in the SMS polling, who are likely to be more interested in consumer issues
- The A to E rating scale used in the question, as discussed above
- The impact of weighting

There were also disparities between the profile of main banks used by respondents, and by the public as a whole.

On the **January 2003** survey respondents were asked about the Measles Mumps Rubella (MMR) vaccine, the findings from which were subsequently benchmarked using the MORI Telephone Omnibus. A comparison of findings between the two surveys is inconclusive; overall, Watchdog respondents appear more sceptical than the public as a whole (although this is not consistent across all age groups): 51% say MMR is safe, versus 64% in the MORI benchmark survey. Whether this is the impact of the sample, the methodology or the weighting is inconclusive.

10. Applications

SMS polling is an emerging technology, facing constraints including:

- The profile of respondents willing and able to take part
- The type and length of question that can be asked
- The amount of information that can be collected to weight the data.

In the future these may be overcome as:

- The use of messaging increases amongst older people

- Messaging technology improves, allowing longer messages with pictures and sound etc.
- Alternative billing structures are introduced to allow reverse billing, removing the financial burden on the respondent
- And a wider range of sample sources become available.

It is likely that there will be a convergence between SMS, WAP and other on-line data technologies that will allow a wider range of questions to be asked (similar to internet surveys).

Regardless of the technology adopted, mobile polling (including all of these technologies) does offer advantages over other survey methods[1]. In particular, the speed of response is unique and provides the opportunity to conduct rapid response research on urgent issues more cheaply than a single issue 'quick turnaround' telephone poll.

Potential applications will be limited by the availability of sample and the extent to which they represent the population as a whole, but opportunities include research amongst specific audiences where a panel can be readily generated, e.g.:

- Customers (who could be surveyed on satisfaction with service once an order has been fulfilled or after a conversation with a call centre)
- Employees
- Young people (subject to parental consent)

Appendix 1

Table 1. Final Sample Profile versus Population

		GB Adult Population Profile	Watchdog Viewer Profile[2]	Watchdog SMS Respondent Profile
Age	15-24	15%	7%	14%
	25-34	19%	12%	34%
	35-44	18%	14%	22%
	45-54	16%	16%	13%
	55-64	13%	19%	13%
	65+	19%	33%	3%♣
Gender	Male	49%	46%	48%
	Female	51%	54%	51%
Class	AB	24%	17%	23%
	C1	27%	29%	41%
	C2	21%	21%	19%
	DE	28%	32%	12%
Work status	Full-time	45%	-	63%
	All excl ft	55%	-	37%
Cars in HH	0	21%	-	14%
	1	43%	-	42%

[1] Earlier identified in the ESOMAR paper *Using WAP Phones to Conduct Market Research: A Report on Preliminary Trials* (Matthews, Bennet and Down, Novemeber 2000)

[2] 16+. Source: BARB, September 2002 to April 2003

		GB Adult Population Profile	Watchdog Viewer Profile[2]	Watchdog SMS Respondent Profile
	2	26%	-	33%
	3+	9%	-	10%

♣ Note, before panel was boosted those aged 65+ made up less than 1% of the panel

Appendix 2: SMS Questions

December 2002: Satisfaction with Banking Services

1st Question:

Watchdog Q:How satisfied or dissatisfied are u with service from your main bank?Pls vote on scale A to E where A=very satisfied->E=very dissatisfied, F=don't know

2nd Question:

Watchdog: Which is your main bank? Please reply using following code, B=Barclays, H=HSBC, L=LloydsTSB, N=Nat West, O=Other bank

January 2003: MMR Vaccination

Preamble:

Watchdog: Next poll is about MMR triple vaccine (Measles Mumps Rubella) given to children

1st Question:

Watchdog: Do you think it's safe or unsafe for parents to choose the MMR vaccine for their children? Reply A=SAFE, B=UNSAFE, C= DON'T KNOW

2nd Question:

W'dog: If Tony Blair said his son was given MMR vaccine, would this make parents MORE (D) or LESS (E) likely to use it, or make NO DIFFERENCE (F). DON'T KNOW=G

Appendix 3: Survey Findings

December 2002: Satisfaction with Banking Services

Table 1.

Satisfaction with the service from your main bank	**SMS Survey Findings**	**MORI Benchmark**[†]
Very satisfied	28%	50%
Fairly satisfied	27%	40%
Neither/nor	26%	5%
Fairly dissatisfied	10%	3%
Very dissatisfied	8%	1%
Don't know	1%	1%
Base	(1,567)	(12,000)

Table 2.

Which is your main bank?	**SMS Survey Findings**	**MORI Benchmark†**
Barclays	10%	17%
HSBC	19%	13%
LloydsTSB	16%	23%
Natwest	16%	15%
Other	39%	26%
Base	(1,302)	(12,000)

† MORI Financial Services, for six months ending June 2002. Both surveys based on weighted data

January 2003: MMR Vaccination

Table 3.

Do you think it is safe or unsafe for parents to choose the MMR vaccine for their children?	**SMS Survey Findings**	**MORI Telephone Omnibus Findings†**
Safe	51%	64%
Unsafe	23%	14%
Don't know/not stated	26%	21%
Base	(1,814)	(1,008)

Table 4.

If Tony Blair said his son was given the MMR vaccine, would this make parents more or less likely to use it, or make no difference?	**SMS Survey Findings**	**MORI Telephone Omnibus Findings†**
More likely	37%	35%
Less likely	5%	1%
Make no difference	46%	60%
Don't know/not stated	11%	4%
Base	(1,683)	(1,008)

† MORI Telephone Omnibus, 31 January – 2 February 2003. Both surveys based on weighted data

About the Authors

Joel Down is Head of Digital Media at MORI, directing research using a wide range of on-line and off-line, quantitative and qualitative methodologies. He works with a spread of public and private sector clients such as Nokia, the BBC, British Airways, the Office for National Statistics and the Department for Work and Pensions. Joel can be contacted at MORI, 79-81 Borough Road, London SE1 1FY; Tel 020 7347 3000, fax 020 7347 3802, joel.down@mori.com

Simon Duke is an Associate Director at MORI, responsible for a range of areas including iTV and MORI's regular Technology Tracker research. He can be contacted at MORI, 79-81 Borough Road, London SE1 1FY; Tel 020 7347 3000, fax 020 7347 3802, simon.duke@mori.com.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Data Collection Initiatives and Business Surveys in the Office for National Statistics

Peter Thomas

Abstract

Over the last seven years the United Kingdom Office for National Statistics has been implementing a series of initiatives to improve the collection and processing of business statistics data in the UK. The paper describes the recent history and covers proposals, which are at present either at a pilot stage or projected for the next four years. A range of new technology solutions have been applied to data collection. Document imaging and scanned forms have replaced paper forms for all processes. For many of the smaller inquiries Telephone Data Entry (TDE) is increasingly being used to collect the data. Pilots of data collection through the Internet have also been carried out. Having virtually all incoming data in electronic format has allowed the introduction of workflow systems across a wide range of data collection activities. The paper covers the future strategy which will primarily centre around the development of data collection on the Internet and changes to the scanning systems that will allow data to be captured by question providing the potential for much greater flexibility in business form design. The paper also covers some of the major processing developments that have focused on automatic and prioritised editing.

1. Background

The Office for National Statistics (ONS) conducts a range of business surveys which form the basis for economic statistics used by the Government in its monitoring of the United Kingdom economy. These surveys are a combination of short term - monthly and quarterly - surveys, often to identify turning points in the national economy, and annual surveys that generally measure levels. Examples of these surveys are:

- Monthly Wages and Salaries Survey used as the basis for the calculation of the Average Earnings Index
- Monthly Production Inquiry used as the basis for the Index of Production
- Monthly Inquiry into the Distribution and Services Sector used as the basis for the development of an Index of Services
- Quarterly Capital Expenditure Survey to measure levels of capital investment in the economy
- Annual Business Inquiry (employment) used as the basis for employment estimates
- Annual Business Inquiry (financial) used as the basis for a range of economic aggregates that feed through into the National Accounts
- New Earnings Survey used to assess earnings by occupation, gender and location.

There are about 100 business surveys each year, the majority of them conducted under the Statistics of Trade Act 1947. Currently 1.8 million forms are sent out each year to more than 320,000 businesses.

Most business survey samples are selected from the InterDepartmental Business Register (IDBR), a comprehensive register of businesses and their structures. The major sources of updating of the IDBR are from Value Added Tax and Pay as You Earn records that identify new and closed businesses, and from the Annual Register Inquiry carried out by ONS to gather local workplace details for multi-site businesses. Sample selections are usually taken from IDBR populations stratified by industrial activity and by size measured by employment. The chance of inclusion in a survey increases with the size of a business. The largest businesses are always included in relevant samples because of their importance to the economic statistics. To ensure that any business survey results are representative, businesses of all sizes and types of business activity have to be included in the survey samples.

2. Overview of the Business Survey Data Collection Innovation Programme

ONS has had a Data Collection Innovation (DCI) Programme in place since the mid-1990s in support of an overall data collection strategy for business surveys. This strategy has three main aims in relation to business surveys:

- To reduce the form filling load on businesses contributing to the surveys
- To improve the quality and timeliness of the business data provided
- To reduce the cost and increase the efficiency of business data collection

At present the main operational elements of the business survey data collection process are:

- Paper-based systems for all surveys, using image scanning and intelligent character recognition (ICR), which has very largely eliminated paper handling beyond the scanning phase
- "Touch-tone" telephone data entry (TDE) for a number of short term inquiries where the amount of data collected on each form is small
- Workflow applications in Lotus Notes to manage the data processing and validation tasks
- Central receipt and indexing of faxed returns which are then redirected to data validation, as opposed to data capture, staff for online data take on.
- A Contributor Comments Database (CCD) on which any relevant information gained from contributors when they are contacted, for example to query their returned data, is recorded

The data collection programme has made significant progress over the years on all three of the aims above. There has been a reduction in compliance as the CCD has lessened the need to contact contributors to query their data. Progress has been made on the more timely provision of data through the use of TDE which is now the collection method used by businesses for over 20% of business forms. TDE has also had an impact on quality through the ability to have some limited validation built into the system at the point of data collection.

The DCI Programme has helped to achieve significant improvements in survey processing efficiency. In this context the use of scanning and ICR, in combination with organisational changes, has led to reductions in excess of 50% in costs. These organisational changes have progressively moved from a structure where statistician led commands managed all aspects of a survey from capture to results, to one where while the results and data collection functions have been separated. Virtually all data collection for business surveys is now concentrated in a single Division, with other Divisions responsible for results. Finally two successful pilots of collection of business data over the Internet have been completed, and further work in this area will be taken forward as part of an overall Modernisation Programme for the ONS.

Over the next few years the main priorities of the next phase of the DCI Programme will be to reduce the burden on contributing businesses and to improve the quality of the data provided, including its timeliness. The Programme seeks to achieve this through the following vision:

- Maintain-paper based systems for each business inquiry using scanning and ICR for data capture. While the volume of paper based returns is expected to reduce, a significant number of businesses are expected to want to continue to provide their data through this medium
- Further take up of TDE in surveys where it is already in place and roll out to other appropriate business surveys
- Develop and implement Internet data collection systems
- Integrate data capture of faxed returns into the scanning and ICR processes.

The business benefits of this DCI vision can be summarised as:

- There will be timeliness benefits where businesses switch to the electronic options of TDE and the Internet. For example, in some of the monthly surveys up to 6% of forms are returned within the two days after the survey closure date
- There will be quality improvements through the ability to carry out data validation at the point of data capture for both TDE and the Internet.
- Data validation at source will also reduce business costs as there will be less need to contact contributors to query their returned data
- Web based data collection will provide a secure means for ONS to receive data from the increasing number of contributors who want to return their data through use of Email
- There will be reduced data collection costs through both a reduction in printing and postage costs, reduced data capture costs where it is returned electronically, and reduced validation costs with more validation carried out at source.

There is no reason why the business benefits outlined above could not be achieved with systems specifically developed for each of the three main data collection mediums - paper, TDE and the Internet. However, the current DCI Programme goes beyond this by moving from the current form/page based recognition, which is very inflexible and resource intensive, to question based recognition. The ultimate aim is to automate the process of business “form” production irrespective of whether the form is paper, telephone or web-based.

The move to question based recognition also has other significant benefits. It would support the development of customised or bespoke forms which would provide much greater flexibility in their design. This opens up the potential to markedly reduce compliance costs to businesses. It would provide a ready facility to incorporate one-off questions into business inquiries and also the ability to enable capture from faxed returns to be integrated with that for paper forms.

The following sections of this report provide more details on the evolution and plans for business data processing within the ONS. It does this in four sections.

Section 3 *Data Capture from Paper Forms (scanning and ICR)*

Section 4 *Telephone Data Entry*

Section 5 *Internet Data Collection*

Section 6 *Integrating data collection across paper, TDE and the Internet*

The above focuses on improvements to the business survey processes by which businessses are able to return there data to the ONS. Section 7 briefly refers two processes, automatic and selective editing, that have been developed to reduce the amount of editing of the returned data, but without impact on the quality of the survey results.

3. Data Capture from Paper Forms (scanning and ICR)

Document Imaging

The first document imaging project was set up in 1995 and tested on the Monthly Turnover Inquiry. The successful pilot allowed document imaging to be rolled out to the remainder of ONS business inquiries. This has now been applied to 95% of business forms received in the ONS.

Kodak Scanners with OCR_for_Forms software are used to take images of the business survey forms. These images are then used to enable the form-type to be recognised, through a form identity number, and then to capture the hand-written data. As part of this process data that cannot be read is flagged and this is then routed to data perfection staff, along with the image, using UNIBASE software. This system operated at a data character recognition rate of around 97%.

In 1998/99 “Drop-Out Colour” technology was implemented, whereby coloured data entry boxes drop out at the form scanning stage to allow the software to recognise only the new data added to the form by the contributor. This has increased the data character recognition rate to approaching 99%, with less resource required for manual verification. All forms, with the exception of a small number of long complex forms with a great deal of free format responses, are now processed by the system. The system has proved to be reliable and relatively easy to use. Following the introduction of drop-out colour a move to standardise on white paper for business inquiry forms has further improved the quality of scanned images and the data capture from returned forms.

The bulk of incoming forms returned by fax now arrive at a central fax server. A simple indexing system enables the electronic fax to be stored along with the images of returned paper forms. The appropriate inquiry processing section is then electronically sent the form, and currently these staff key the data from the fax image.

Use of Lotus Notes

The organisational changes in data collection and the move towards electronic versions of data, whether document images, TDE or Internet, gave rise to demands and opportunities for improvements to working procedures. Lotus Notes was selected as the medium for the introduction of new workflow systems. Apart from the widespread adoption generally for this kind of application, Notes is also used by the other national statistical institutes providing a ready made pool of relevant experience which the ONS could share. The initial pilot systems proved very popular with the users and rapid growth in the user community and the range of applications proceeded faster than our ability to manage the expansion. The strengths of the Notes applications were that they reinforced the move away from single inquiry workgroups and allowed processes, which were similar for inquiries, to be dealt with in a consistent way.

A major success has been the Contributor Comments database, which allows all inquiries to share the soft information arising from comments on forms, conversations with contributors or desk research into business organisation. This has reduced the effort required to explain unusual data and eliminated duplicate telephone calls from different inquiry areas to a single contributor.

Notes is also an excellent tool for work allocation, prioritisation and monitoring, which has eased the management of work areas with responsibility for a mix of short term and annual inquiries. Some of the work arising from a form, such as change of address or business structure, is more appropriate for

the business register than the inquiry and the combination of Notes and electronic sources of data makes it easy to reassign work in a secure and audited environment.

The problems with the initial Notes applications were partly the problems of managing the rapid expansion, but a major technical drawback was the difficulty of communication between Notes and the legacy inquiry systems, which remained vital for the number crunching processes such as validation and analysis. Systems became hybrid, with users having to move between old and new systems and some bulk data being transferred to Notes where it was no longer being updated in real time.

Proposals for Document Imaging

Document imaging is now being developed in the second stage of the DCI Strategy. Over the next two years proposals will include the improvement of the Intelligent Character Recognition process, with a move to question-based recognition rather than the current form-based recognition. Data capture will move to being by individual question rather than for a set of questions fixed by the template for each page of an inquiry form, thus giving greater flexibility in the design of the questionnaire and reduced maintenance costs. The new methods will also allow the procedures used for handling paper forms to share common systems and standards with other media, via the Collection Database which is discussed later.

Improvements in the processing of faxed data are also planned. The new question-based recognition software currently under development will allow the integration of the output from the central fax server with the ICR process used for paper forms. This will enable automated data capture straight into the inquiry processing systems, and removes the need for online data take-on by the inquiry staff.

4. TDE “Touchtone” telephone data entry (TDE)

Data from forms with only small numbers of variables are increasingly entered via telephone data entry technology. In 1995 the first pilot of telephone data entry was carried out for the collection of data for a new Service Sector Price Index. Contributors returned price information on a range of products. New contributors were offered telephone data entry as the means for supplying data for the inquiry.

Telephone Data Entry uses the tones of a telephone keypad to make responses and to allow the contributor to undertake a dialogue with a set of recorded messages. Contributors may optionally add voice messages to explain validation or credibility problems.

Contributors received a routine hard copy contact letter each period requiring them to provide the price for a stated product or products. With TDE, this requires a contributor to enter a reference number and then a price for the product(s). There is also a facility for a contributor to leave a voice message, which can be played back by the survey processing staff. The system allows for data entered to be automatically checked against a previous return, including a prompt for a comment from a contributor where new data entries are inconsistent with previous data, thus providing a vital validation check.

TDE was extended to the Producer Price Index, where the majority of price quotes per month are supplied this way. There is an option in the collection of data for the Retail Sales Inquiry, which enables contributors to choose whether to supply data on the paper form or using TDE. The facility to use TDE is now offered on over 40% of short-term inquiries. A recently launched Vacancies Inquiry has virtually 100% of its data capture via TDE.

Initially there were problems with capacity, in dealing with comments provided by contributors, and to a lesser extent within the telephone network. These were all been ironed out and the system was very reliable and continued to run successfully over the "Year2k" period without intervention. However, the popularity of TDE as a data entry mechanism created its own problems within the ONS because the original system had only a limited number of contact telephone numbers. The hardware of the system was therefore reviewed to increase the overall capacity of the system, and to provide additional facilities.

The original system ran on non-standard hardware using a non-standard operating system with analogue telephone lines. Since the original system was purchased the market and technology had moved on considerably. The system was therefore replaced, which met the following objectives:

- move to a Windows NT platform;
- expansion of system capacity;
- resilience of the system;
- increase in the capability of the system;
- provision of desktop capability, including inter-connectivity with ONS' telephone handsets (although this is not yet in use);
- provision and use of ISDN (digital) capability;
- reduction in the maintenance cost per line of the system, although it is recognised that an increase in capability may mean an increase in maintenance.

Future Proposals for Telephone Data Entry

Now that the system has been upgraded the inquiries using the system are being expanded. Telephone data capture through TDE now accounts for over 20% of returned business survey forms. A roll-out programme is underway to offer TDE on other suitable business inquiries, with expectations that the percentage of forms returned by TDE will increase to 30% . This roll-out is now using dialogues that have been standardised. TDE is also used to provide other facilities, for example to order a duplicate form, and further similar applications are under consideration. The use of voice recognition will also be considered.

Investigation will be made into the business requirement for Computer Assisted Telephone Interviewing (CATI) and Computer Telephony Integration (CTI). For example, applications could include assisted dialling for response chasing or a purely telephone-based inquiry.

5. Internet Data Collection

Data collection via the Internet was piloted between 2000 and 2002 on two business inquiries to ascertain whether the concept of Internet data collection would be successful for business inquiries. This work was stopped in 2002 as it was realised that the technology to develop the next generation of Internet data collection systems within ONS business surveys would be different from that used for the pilots. Proposals for the future work on Internet data collection for business surveys are expected to be taken forward as part of ONS's Modernisation Programme. A key focus of this Programme is the standardisation and systemisation of processes used within ONS for business surveys, and also across social and administrative data capture where appropriate.

The purposes of the two pilots were similar in nature in that as well as proving the technical feasibility of Internet they would also provide an assessment of the potential and benefits, including efficiency

savings, from the use of Internet for data collection. The lessons learned from these pilots would enable recommendations to be made that would guide the further development and implementation of Internet data collection. Finally they would enable ONS to make progress towards E-Government targets of all business surveys having electronic options by which businesses could return their data.

The development process for the pilots was similar starting with contact with contributors to establish how many would be prepared to take part in the pilot. This was followed by the development of prototype systems, which were subject to amendment arising initially from the testing process, and then from feedback from users of the system and contributors in the pilots. Guidelines for use by both the contributors and the staff that would process the returns were compiled and detailed training given to the processing staff. Prior to the pilots going live, agreement was reached on the information needed to support their evaluation, including whether there were any modal effects on the data. These were based on control groups of contributors who had similar characteristics to those in the pilots. The main difference between the two pilots was that the second one to be set up included a registration system as part of the arrangements to manage the security of the data.

The points below summarise the main lessons learned during the pilots:

- Not all contributors have internet facilities and requested paper forms;
- There were still issues of non-response and paper forms were required for enforcement purposes;
- The “feedback” database was successful as a way of gathering contributors views;
- The web was quick and user friendly;
- Problems identified by contributors were accessing the system using their user ID and password, browser problems and server availability;
- The Registration process created delays in response;
- The on-line validation in one of the pilots was thought excessive by the contributors;
- No print facility for contributors for previous and current returns;
- Problems when contributors closed down system and tried re-entering to amend data.

Information on the level of take-up of Internet as a data collection option by businesses comes from both of the pilots. In the first a decision was made to increase the number of Internet contributors. This proved to be a time consuming exercise with 130 contacts made with contributors in order to get a further 39 involved in the pilot.

The second pilot provides information on the extent to which businesses will use the Internet once they have expressed an interest in providing their data by this method. Of the 126 potential Internet contributors,102 (81%) provided some feedback over the Internet with 83 (66%) providing acceptable Internet returns. For the others there were problems with difficulty or inability to register and some not in fact having Internet access.

In terms of the staff processing the Internet returns the view was that after a period in which a contributor had to settle into the use of the Internet in providing their data they preferred processing data collected via the Internet. While it is recognised that the pilots were quite small in scale, investigation found no evidence of modal effects for data supplied over the Internet. Internet also provided a direct benefit to contributors in that a “reward” system was built in with contributors able to access screens that allowed them to compare information for their business with trends across the industry as a whole.

As part of ONSs modernisation Programme, future DCI work, including that on the internet of the collection tool is expected to:

- Recognise and incorporate the ONS corporate dimension and work as part of corporate projects for technical development of all aspects of the DCI programme including the Internet;.
- Incorporate appropriate registration and authentication;
- Replace returns of unstructured data by a "secure" E-mail system.
- Follow design standards for data collection instruments;
- Take on board "lessons learned" and recommendations from the pilots.
- Work with JAVA and Oracle the corporate tools, to collect data into the Corporate ONS Repository for Data (CORD), including integration with CORD irrespective of the means by which data is collected. The Business Collection Database referred to below is expected to be integrated within CORD.

As it currently stands, the overall DCI programme and of which Internet data collection is a major part, is expected to reduce business survey processing costs by £800K (some 10%).

6. Integrating data collection across paper, TDE and the Internet

In the future it is planned that data for business surveys will be returned by paper form, fax, TDE, or by Internet. Developing and implementing these as separate systems implies large overhead, maintenance and data integration problems. We expect to solve these problems with the introduction of the Business Collection Database.

The Collection Database will be used to drive all the processes connected with the capture of data for inquiries, integrating the various collection media. It will underpin the collection of data, irrespective of the medium by which that data is collected - it will cater for data returned to the office on paper, by fax, via the Telephone Data Entry (TDE) system, and via the Internet. Eventually the database will also drive the production of the collection instrument; ultimately it could potentially be used by the Data Validation Unit when validating and editing returned data.

Benefits of this approach include:

- a single integrated system supporting and driving data capture via all types of media, and managing increasing mixed mode data collection in a standard way;
- increased commonality and standardisation of processes, and therefore reduced maintenance costs;
- the provision of considerable data for the management of the collection processes, including for audit and monitoring processes;
- one of the essential foundations that must be in place before a "Bespoke Forms" production system can be implemented for the fast and flexible production of forms; which leads to increased flexibility for customers to change inquiry questions and the ability to vary questions within inquiries and a reduction in compliance costs;

7. Methodological work in support of business data capture

In spite of attempts to use best practice design questionnaire techniques to minimise the number of errors, respondents still make mistakes. These can be as simple as returning data in pounds instead of in thousands of pounds, or more complicated through misunderstanding the questionnaire or the inability to provide data as requested.

Validation techniques are built into most of the ONS survey systems to identify such errors and also to identify suspect data. Predetermined validation gates are used to identify either large changes in data items. This can be by comparing individual data items with those returned for previous periods, or for new respondents, comparing their data with those expected for a particular industrial classification or employment sizeband. The validation system will also identify missing data or errors, such as incorrect dates, totalling errors. Some surveys also employ congruency checks comparing similar data collected from administrative sources, such as held on the business register , and from other surveys, such as 12 months turnover from a short-term survey with that collected for an annual survey.

Validation consumes a substantial proportion of survey resources - generally thought to account for up to 40% of total survey costs. Traditionally, National Statistical Institutes have believed that focusing a large proportion of resource on data editing and validation produces high quality survey data. However, over the last decade the old philosophy has viewed as inefficient, leading to high survey costs, high respondent burden and possibly poor quality data due to errors being introduced into the data through over-editing. The new philosophy advocates that good data quality is not guaranteed through large amounts of data editing and that valuable resources should only concentrate on the "important" suspect values.

The ONS has been exploring the introduction of new editing techniques such as selective or priority editing. The project concentrated on short-term improvements to existing systems, rather than replacing them. The aim was to develop and apply two new methodological approaches to improve the efficiency of data validation and editing without adversely effecting data quality.

Automatic correction of systematic errors

Satisfactory methods of automatic correction were developed for two kinds of systematic errors:

- thousand pounds errors, in which respondents give information in pounds when asked to provide it in thousands of pounds;
- totalling errors, when there is inconsistency between a reported total and the sum of its reported components.

The automatic editing method for thousand pounds errors focuses on the main variable value returned at respondent level, in this case, total turnover. The method operates by comparing the returned value in the current reporting period with the accepted value from the previous reporting period. Where the ratio of these two values falls within a specified range centred on 1000, the returned value is adjusted by dividing by 1000 and rounding the value. The automatically edited value is then used in all further data processing.

The automatic totalling method was introduced for employee totals. It focuses on the total employee value, returned at the respondent level for business surveys and at the local unit level for the business register. The method compares the sum of the components for the current reporting period with the accepted total employment value from the previous period. If the difference lies within an acceptable range, then total employment for the current reporting period is amended to equal the sum of its components. The automatically edited value is then used in all further data processing.

Both methods of automatic editing operate before the data are passed through the survey specific validation system, hence reducing the errors that are triggered by the validation system. This in turn demands less editing resource and ensures that this type of error is edited less subjectively. These methods were successfully piloted on several live survey processing systems and have been introduced to all suitable surveys.

Selective editing

This prioritises validation failures so that only those for which correction is expected to have a material impact on the survey outputs are followed up for editing**.** This process takes place as part of batch data take-on and follows automatic editing. Data changes arising from many validation failures result in negligible changes on the survey estimates. Selective editing allows for individual businesses data that fails standard system validation checks to be assessed and scored according to its level of impact on the survey output.

The possible application of selective editing to business surveys in ONS was investigated and analysed, using four months of data from one key ONS business survey, by the University of Southampton (UoS). The UoS methodological evaluation study identified the key variables that required scoring according to its level of impact on the survey output. Individual businesses data with scores above a predetermined threshold were identified as requiring scrutiny and editing, and those below the threshold were accepted without further scrutiny. The pilot of selective editing on one ONS business survey resulted in a reduction in the number of data items requiring scrutiny of over a third. The effects on the survey results were confirmed as negligible.

Selective editing has now been rolled out to all appropriate short term surveys. Although the principle remained the same this did involve identifying specific key variables for the scoring of each survey's data. Studies are underway to investigate the possible application of selective editing to longer term and more complex surveys. This will take time as piloting has demonstrated that the same method cannot be simply transferred to another survey without impacting on the quality of the survey estimates.

The introduction of selective editing led to the development of graphical editing to support selective editing and to provide DVB supervisors with a tool for quality assuring the work of the data analyst. During the pilot of selective editing it was noted that errors in the non-key variables may slip through the selective editing process, as the score only takes account of the key variables. Some additional form of editing was therefore required to check for significant errors in the unedited values for the non-key variables. Graphical editing performs this role by the plotting of graphs by industry, so that the analyst builds up a knowledge of the industry, recognises patterns and relationships and therefore detects anomalies in the data. Supervisors can also focus their quality checks by graphically determining the large changes in the weighted and unweighted contributor data.

8. Conclusion

The ONS has made considerable steps over the last seven years in improving its business data collection processes. Further challenges remain ahead including the introduction of web based data collection and the ability to generate bespoke forms. Much of the change over the next three years will be through the ONS's Modernisation Programme.

About the Author

Peter Thomas works for the Office for National Statistics in the UK and can be contacted at Peter.Thomas@ons.gsi.gov.uk.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Interviewing Children Using Audio-CASI (Computer Assisted Self Interviewing)

Rebecca Gatward

Keywords

Audio-CASI, Usability, CASI, Young people and Mental health

1. Introduction

In March 2001 Social Survey Division (SSD) of the Office for National Statistics (ONS) carried out the first pilot stage of a survey of the development and well-being of children and adolescents looked after by local authorities. Looked after children are either in the care of or accommodated by Local Government Departments. The mainstage of the survey was carried out between January - March 2002. The survey questionnaire includes a substantial set of sensitive questions asked using audio-computer assisted self-interviewing (CASI), the decision to use audio-CASI is explained in more detail later in the 'Background to the survey' section of this paper. This was the first time SSD had used audio-CASI as a method of data collection.

Gaining any young person's co-operation in a survey can be problematic; however, looked after children are more likely than other children to have conduct disorders, emotional problems or hyperactivity. So they form particularly challenging subjects.

This paper will focus on usability and the practical aspects of using audio-CASI with this group of young people. Information for the paper was obtained from interviewers and respondents. Respondents were asked to answer a short set of questions about audio-CASI at the end of their interview. Feedback from interviewers was collected at a face to face debrief session. The paper also describes some of the more practical aspect of developing an audio-CASI instrument and would be of particular interest to others who are planning to use audio-CASI for the first time. The paper also assesses the benefits of using audio-CASI as a mode of interviewing amongst this particular group of young people. Conclusions made in this paper could also be related to other challenging groups of respondents. Other aspects that may also be of interest are that we used standard laptops and that the audio-CASI instrument included some open text questions.

Audio-CASI has been used extensively as a mode of collecting data on sensitive behaviours and this is obviously not the first time the method has been used when interviewing young people. During the development of the audio-CASI instrument we drew on the work of others who have used audio-CASI particularly the work of Jim O'Reilly.

2. Background to the survey

In 1999, ONS carried out the first national survey of the mental health of children and adolescents in private households in Great Britain (commissioned by the Department of Health, the Scottish Office and the Welsh Office). Interviews were completed about 10,438 children and adolescents aged 5-15 years. This study showed that 1 in 10 children aged 5-15 had a clinically recognisable mental disorder: anxiety, depression, hyperactivity or behavioural problems which had a severe impact on the family. It also showed that children with a mental disorder were three times more likely to have a specific learning difficulty than those without a disorder. In this previous survey the self-completion questionnaire was administered using CASI. For this new survey the questionnaire was developed as an audio-CASI instrument.

Audio-CASI was used as the mode of data collection for two main reasons. Firstly, the sensitive nature of the questions. It is well documented that audio-CASI can increase the reporting of sensitive behaviours. For example Turner et al (1998) reported that respondents were much more likely to report risky behaviours when they were interviewed with audio-CASI measurement technology than when interviewed with more traditional paper self-administered questionnaires. They also state that audio-CASI appears to have a more pronounced effect on the reporting of behaviours that are particularly sensitive, stigmatised or subject to serious legal sanctions, compared with less sensitive areas of conduct. Although it has been found that CASI and audio-CASI are equally effective at obtaining reports of sensitive behaviours, respondents prefer audio-CASI as a reporting method. (O'Reilly et al, 1994).

The other main reason for using audio-CASI was the nature of the respondent. The young people in this survey sample were more likely to have learning difficulties and problems with concentration. Audio-CASI appears to be a suitable mode for administering surveys to low-literacy respondents, particularly when the survey is collecting personal or sensitive material. (Schneider and Edwards 2000). We hoped that using audio-CASI would enable the young people with learning difficulties to take part fully in the survey and increase the young people's level of attention and interest in the interview.

Not all the young people in this sample had learning difficulties, some had concentration problems and others just wanted to be doing something else or be somewhere else. Therefore it was necessary to develop a questionnaire that could be easily used by the young people who had learning difficulties, without appearing cumbersome or simplistic to respondents who are proficient readers (Couper, 1998).

3. Survey methodology

The survey involved collecting information about each child from up to three sources: parent, child and teacher. The questionnaire was composed of the following five sections administered using various modes and all contained within one datamodel;

- Parent/Carer face to face interview
- Young person face to face interview (11-15 year olds) (P)
- Young person self-completion section (11-15 year olds) (P)
- Specific learning difficulties test (all children 5-15 years)
- Teacher questionnaire (postal questionnaire, keyed into the datamodel when returned to office).

The two sections, marked with 'P', were set up as parallel fields to allow interviewers flexibility over the order in which they were administered. Depending on the number and complexity of problems the young persons face to face interview could be very long. We felt it was important that interviewers had the option to change to a different section if the child started to lose interest.

4. Design of the Audio-CASI questionnaire

The young person self-completion questionnaire included questions on the following topics:

- Moods and feelings
- Troublesome behaviour
- Cigarette smoking
- Drinking (alcohol)
- Experience of drugs
- Sexual activity
- Exclusion from school

The questionnaire included both closed and open questions. When developing the questionnaire we set a limit of four response categories for the closed questions. The open questions required only short text or numeric responses, for example: *What word best describes how you have felt in the past 2 weeks?*, up to 20 characters to be entered at this question.

The audio-CASI instrument was developed from a CASI questionnaire that had been used successfully in a previous survey of children. Some less sensitive questions that were in the CASI version were not included in the audio-CASI version but instead asked in the face to face interview. These were mainly questions that had more than four response categories and were part of standard instruments that could not be changed. When designing the audio-CASI instrument we also tried to avoid frequent changes in sets of response categories and to keep the number of different response sets to a minimum. In total the self-completion questionnaire (approximately 130 questions) included 6 different response sets.

The questionnaire was programmed using Blaise 4.2 and administered using interviewer laptops with the standard set up.

Interviewer assisted audio-CASI

The audio-CASI instrument was interviewer assisted. The section started with an introduction read by the interviewer in which they explained how the section worked i.e. they would hear the questions and possible answers through headphones and then enter their own answers into the laptop. Interviewers demonstrated to the child how to enter their answers, how to move on to the next question and how to repeat a question, the young people were also encouraged to ask for help if they needed it.

We hoped that this step by step approach would discourage the young people from being alienated and give them confidence to complete the section. At each stage the interviewer or the child was told via the headphones what they were required to do or happens next.

Changing mode of completion

Ideally we wanted all young people to complete the self-completion section using audio-CASI. However we were aware of the possibility that some young people would be too anxious or have hearing problems. Three possible modes of completion were available; audio CASI, CASI or CAPI

(questions read by the interviewer). Instructions and screen layouts were programmed for each of the three modes.

Checking the volume level

Before handing the laptop to the child interviewers checked that the headphones were working and adjusted them to a reasonable volume. The interviewers then passed the headphones to the child who then heard the following instruction: *'The first question you will hear will be a test question just to check that the volume is ok for you'*. This was followed by: *'Is the volume OK for you?'*, if the child said 'no' they were then prompted to ask the interviewer to adjust the volume. The question was then repeated to ensure that the volume was at the correct level.

Practise question

Before starting the 'live' questions a simple practise question was asked to check that the respondent understood how to enter their responses.

Screen layout

Our aim, when designing the audio-CASI version of the questionnaire, was to keep the screen uncluttered to avoid distracting the respondent. The question text was not displayed on screen because we felt it was more important for the child to concentrate on listening to the question, we felt this was especially important for young people who had reading difficulties. The response categories were displayed in the answer list section of the infopane in the standard way.

Obviously if the section was completed using CASI or as a face to face interview it was necessary to display the question text. To achieve this the colour of the question text changed according to the mode of completion. If audio-CASI was used the text of the question was set to the same as the screen background and therefore could not be read, if CAI or CASI the question text was set to a contrasting colour so it could be read.

The colour switch variable was set up as a parameter in the top-level block of the self-completion section of the questionnaire. Appropriate colours were assigned to 'W' and 'P' in the modelib editor. The parameter (PColour1 and PColour2 – see example below) was switched between '@W'or '@P' depending on the mode of completion.

C3G3 "^PColour1 Have you ever used GLUE, GAS OR SOLVENTS?@/
PRESS 1 for NO, 2 for Yes @/
PRESS the WHITE key TO CONTINUE@/ ^PColour1
^PColour2@/ Question G3 ^PColour2@/"
MML "SOUND(C3G3.WAV)
SOUND(Delay.WAV)
SOUND(NoYes.WAV)
SOUND(White.WAV)
: NY

Even if the section was completed using audio-CASI it was still necessary for interviewers to be able to recognise the respondents whereabouts in the questionnaire, if they required help. A small identifier was displayed. We were careful not to use an identifier that would be distracting to the respondent such as, the question number or question 1 of 250, so we used the questionnaire variable name.

Labelled keys

The main navigational keys were colour coded, using paper stickers. The <ENTER> key was labelled white and the <F10> key, which had been assigned as the repeat key, was blue. Respondents were also given a card with the following three simple instructions;

use the white key to get to the next question

use the blue key to repeat the question

ask the interviewer if you need any help

When choosing which colours to assign to the two navigational keys we avoided colours that would be confused by people who were colour blind. We also used good quality stickers that would stay attached to the keys.

Audio guidance provided to respondents

Respondents were given audio instructions about how to record their response to a question and to proceed to the next question. For example, if the response categories were simply 'no' and 'yes', they would hear the instruction *'Press 1 for no and 2 for yes'* and then *'Press the white key to continue'*.

Our aim was to maintain a balance between firstly, providing sufficient guidance and secondly, avoiding frustration for the child by repeating the instructions too often. Instructions were displayed on the first few questions in the questionnaire and then repeated again after each change in response set. If the respondent was expected to record a different type of response such as, typing in their age or text responses they were provided with further instructions. For example, '*Please type in your age in years and then press the white key to continue'*. At these more complicated questions respondents were also reminded to ask the interviewer if they needed help. Other instructions that referred to specific questions were also included.

5. Recording the audio files

When deciding on the best voice for the audio files, and how to ensure a good quality recording we followed recommendations by O'Reilly (O'Reilly, I998). A colleague at ONS was chosen as our voice. This meant that they would be easily available to record any additional audio files or make any essential amendments. The person we chose had a neutral and clear voice and did not have a strong regional accent. She also had a great deal of interviewing experience and was familiar with the questionnaire.

The audio files were recorded at a professional recording studio. This meant that the questionnaire text had to be finalised earlier in the questionnaire development stage than usual. In this instance it did not create any difficulties because the majority of the audio-CASI questionnaire was a replication of a CASI questionnaire used in a previous survey. The questions, response sets and instructions were all recorded as separate .WAV files.

Using the recording studio was an efficient method of recording the large number of audio files required (140). The recording process took approximately 90 minutes and cost approximately £200 (about $300) which included recording time, editing and CD writing.

6. Other practicalities

Other practical issues considered were:

- Headphones - we purchased reasonably priced, robust and compact headphones.
- Screen wipes - interviewers were provided with screen wipes to use after each interview. These were intended to be used if the young person touched the screen.
- Saving the partially completed interview before passing it on to the child for the audio-CASI section – interviewers were given an on screen reminder to save their interview before passing the laptop over to the child.

7. Response and length of interview

All twenty-one 11-15 year olds chose to complete the self-completion using audio-CASI and none of the young people who started the audio-CASI section gave up part way through.

The audio section took, on average 25 minutes to complete. The total interview with a young person, including the face to face interview and learning difficulty tests, took on average 110 minutes.

8. Feedback from respondents

Feedback was collected from the young people via a short set of questions which were administered by the interviewer at the end of the audio-CASI section (a copy of the questions is included as an appendix to this paper). All 21 young people, who completed the audio section, provided feedback.

Overall, feedback from respondents was very positive, all the young people seemed to enjoy completing the audio-CASI questionnaire.

Previous experience of computers

All but one of the young people were frequent users of computers, and all had some previous experience.

General difficulties using audio-CASI

A third of the respondents 'got stuck' whilst completing the questionnaire. In only three instances were the problems related to audio-CASI. The remainder were difficulties interpreting the questions. Problems that did relate to audio-CASI were:

> *'Not much – just the typing',*
>
> *'I think it's because the computer said press enter and I didn't know how to go back',*
>
> *'May be when I heard the question I heard it wrong, and forgot what it was, I was listening to the question I was thinking about the answer but forgot what the question was, I think it would have been better if I had listened more carefully'.*

Voice/Volume

The majority of respondents (15) stated that they could hear the questions all of the time, five could hear them most of the time and just one said they could only hear them some of the time. When asked about how well they could understand the person asking the questions, most of the respondents (16)

said that they could always understand the person asking the questions, three could hear them most of the time and just two said they could only understand some of the time.

Instructions

Only two respondents felt that the instructions were difficult to follow, the remainder felt they were easy to follow or about right. Respondents were also asked about the frequency of the instructions played after the questions. Four felt they were repeated too often, the majority, 12 respondents, thought they were played at about the right frequency and five respondents thought they were not repeated enough.

General comments about audio-CASI

Respondents were asked to say in their own words how they found their experience of audio-CASI. The most frequent response was 'easy', which was given by ten respondents. Five thought it was 'ok' or 'alright', two had 'no problems', one felt it was 'too easy'. This was despite the fact that a third of respondents had previously stated that they had 'got stuck' at some point during the questionnaire, this implies that the young people felt that overall they had not had any problems.

Respondents were given the opportunity to say whether they had any other problems that they had not already told us about – none of the young people said they had any.

Finally, respondents were asked whether there was anything else they would like to say. Twelve respondents did not add anything. The remaining responses were all positive remarks about aspects of audio-CASI or audio-CASI overall.

'it were brill',

'apart from I would like to do it again',

'good it was good fun',

'I wish I had one',

'it's fun',

'I think it is a wicked idea that a computer can speak to you, it is so clever',

'I liked using it, I love laptops',

'it's good'.

9. Feedback from interviewers

Feedback from interviewers was also very encouraging. All the interviewers agreed that the audio-CASI section was the most concentrated and well-received part of the interview. Interviewers got the impression that the young people really enjoyed completing this section. Often, respondents who were distracted during the face to face interview were really interested and did not fidget or show any signs of frustration whilst completing the audio-CASI section. Interviewers felt that one of the reasons why this section worked so well was because the question text was straightforward and therefore easy to understand.

Interviewers reported that the young people interviewed for this survey did generally have lower concentration levels or were more easily distracted than those they had interviewed for a previous survey of young people. During the face to face interview, which preceded the self-completion section, interviewers sometimes found it difficult to persuade the young people to keep going or had to work

hard at maintaining their interest. In some cases interviewers used the audio-CASI section, and the opportunity of using the laptop, as an incentive to keep them going. Interviewers also found it helpful to jump between the different sections of the questionnaire (via the parallel fields).

Interviewers also reported that the whole process of introducing the audio-CASI section to the respondent worked well. Once the interviewer had explained what they were required to do the respondents just '*got on with it*'. There were very few instances when interviewers had to provide assistance to the young person during the completion of the section. Interviewers found that respondents did ask if they required help and if not, interviewers were able to identify when the young person was having difficulties.

According to the interviewers the young people did not have any problems navigating through the questionnaire. One interviewer decided to conduct her own experiment by not labelling the keys on her laptop, instead she told her respondents to press <ENTER> to move on to the next question and the <F10> key to repeat the question again. Her respondents were still easily able to find their way through the questionnaire.

The young people were not tempted to '*play around*' with the laptop once they had come to the end of the audio-CASI section; they just took off their headphones and let the interviewer know as soon as they had finished.

Some respondents seemed to enjoy it so much they asked the interviewers if there was anymore.

Finally, interviewers were asked what they did whilst the young person completed the audio-CASI section. All the interviewers stayed close by in case the young person had difficulties. Some read a newspaper or paperwork, others checked their papers, one interviewer looked at the young person's pets and another talked with the child's carer.

10. Open text questions

There were very few missing responses to the open text questions. In total the questionnaire included 27 open-ended questions, out of 89 responses there was only one missing answer and this was in response to a question asking why they had been in trouble with the police. The quality of some of the open text responses was poor, mostly due to incorrect spellings. However, only one of the responses was unrecognisable.

11. Some lessons for the future

Option to toggle between languages

Whilst developing the questionnaire for the pilot stage we found that it was not possible to toggle between playing the audio files and switiching them off, depending on the mode of completion. As none of the respondents chose to complete the questionnaire using CAPI or CASI at this pilot stage this did not cause difficulties. However, a method needs to be developed that does not rely on interviewers remembering to plug in the headphones in order to stop the audio-files being heard. Ideally we would like to be able to toggle between languages depending on the mode of completion.

The three specific languages would be:

ENG (text) with MML (sound)

ENG (text) without MML (sound)

MML (sound) without ENG (text)

Stop respondents entering early answers

During the pilot respondents were able to interrupt the question by entering early answers. This happened even when the check box 'stop on key' (in the multi-media section of the modelib editor) had not been ticked. If the box is not ticked the respondent should have to listen to the whole question before entering their response (this is not a problem in Blaise version 4.3 and higher).

Checks, signals and error messages

No checks or signals were used in the audio-CASI questionnaire and no guidance was provided to respondents about error messages. Although this did not cause a problem on the small number of cases interviewed for the pilot it is a design issue that needs to be developed further before the mainstage.

Mistake facility

Some guidance needs to be provided to respondents about what to do if they make a mistake. At the pilot stage no instructions were given on how to go back and change their answer. This did cause confusion to one respondent.

Confidentiality

Although lack of confidentiality was not a concern during the pilot we would like to ensure that respondents are not able to go back through the responses the carer/parent gave during their interview. Our solution was to set the section up as a parallel field. This limited navigation to some extent but a more secure method needs to be developed for the mainstage survey.

12. Conclusion

Feedback received from the respondents and interviewers who took part in the pilot suggests that this particular group of young people enjoyed completing the audio-CASI questionnaire and were able to do so independently. Using audio-CASI as the data collection method enabled us to obtain information from respondents who would normally have refused to continue earlier in the interview.

Our main aim was to design an audio-CASI instrument that did not alienate the respondents and provided those with learning difficulties with the confidence to complete the questionnaire: This was achieved by:

- Providing clear guidance via on screen and audio instructions.
- Maintaining a balance between providing sufficient guidance and becoming too repetitive.
- Introducing the section 'step by step'.
- Encouraging the respondent to ask questions.
- Minimising the effort required from the respondents.
- Including questions with straightforward wording.

References

Couper, M.P. (1998), 'The application of cognitive science to computer assisted interviewing. *Cognition and survey research* New York: Wiley, 277-300.

O'Reilly, J.M., et al (1994), 'Audio and video computer assisted self-interviewing: Preliminary test of new technologies for data collection' *Journal of official statistics*, 10, 2, pp. 197-214.

O'Reilly, J.M. (1998), 'Recording audio for audio-CASI instruments'. Unpublished paper

Schneider, S.J. and Edwards, B. (2000), 'Developing usability guidelines for audiocasi respondents with limited literacy skills' *Journal of official statistics*, 16, 3, pp. 255-271.

Turner, C.F., et al (1998), 'Adolescent sexual behaviour, drug use and violence: increased reporting with computer survey technology' *Science*, May 1998.

Appendix A

Feedback questionnaire for young people

IntroF

INTERVIEWER: Please explain that we are interested in finding out what they thought of hearing the questions through the headphones and entering the answers on the laptop themselves.

CompUse

RUNNING PROMPT

Have you used computers....

1. a lot

2. a bit,

3. or have you never used a computer before?

Probs

Did you get stuck at all?

1. Yes

2. No

WhatPrbs

IF PROBS=YES

Where did you get stuck?

PLEASE OBTAIN AS MUCH DETAIL AS POSSIBLE

Hear

RUNNING PROMPT

Could you hear the questions...

1. all of the time

2. most of the time

3. or just some of the time?

Voice

RUNNING PROMPT

Could you understand the person asking the questions...

1. all of the time

2. most of the time

3. or just some of the time?

Instr

RUNNING PROMPT

Were the instructions....

1. easy to follow

2. about right

3. or difficult follow?

InstrRp

RUNNING PROMPT

Were the instructions that are played at the end of some questions repeated....

1. too often,

2. about right

3. or not enough?

KeyB

How did you find entering your answers into the laptop?

INTERVIEWER: PLEASE PROBE IF HAD ANY PROBLEMS

AnyOth

Are there any problems that you have not already told me about?

INTERVIEWER: PLEASE PROBE IF HAD ANY PROBLEMS

AnyCom

Is there anything else you would like to say?

About the Author

Rebecca Gatward works for the Office for National Statistics in the UK and can be contacted at rebecca.gatward@ons.gsi.gov.uk.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Impact of CASI on a Survey Among Gay and Bi-Sexual Men Using Sexual Health Clinics in the UK

Neil Macdonald & Gillian Elam

Abstract

Computer-assisted self-interviews (CASI) have been successfully employed as components of studies of sensitive and risk behaviours. Surveys that depend solely on CASI are a relatively new departure, used mainly in Internet-based studies. This paper discusses the effectiveness of using computer-assisted self-interviews in a clinic-based case-control study investigating sexual behaviour among gay and bisexual men who HIV test (the INSIGHT study).

INSIGHT (Investigation of New Seroconversions In Gay men who HIV Test) was established to investigate the current risk factors for HIV seroconversion in gay and bisexual men, which remain the group at greatest risk of HIV infection in the UK. CASI was chosen as a means of data collection for the case-control component of the study because it shares many of the advantages of a 'face to face' interview (e.g. interaction and validation) whilst preserving the patient's anonymity and standardising interviewer bias.

The CASI program was developed at the Communicable Disease Surveillance Centre (CDSC) using SNAP Version 6 software. The questionnaire is administered using computers located in seven Sexual Health clinics in Manchester, Brighton and London. Patients complete the CASI having received the result of their most recent HIV test and responses are submitted via email to the INSIGHT team at the CDSC. Recruitment began in September 2002 and will continue for two years, collecting data on patients' sexual behaviour and lifestyles during the interval between their last two HIV tests. Data from subjects that test HIV positive (cases) will be compared to that from subjects who remain HIV negative (controls).

A study of this nature poses a number of challenges. The introduction of Computer Survey technology in to Sexual Health clinics and convincing health care workers of the merits of this technology has been a major advance. Implementing the study has placed further demands on services that were already stretched to capacity by increasing numbers of patients.

Self reported sexual behaviour is notoriously difficult to measure. We rely on the patient's ability to remember their sexual practices and risk behaviours between their HIV tests and to feel confident that they can report this behaviour without fear of judgment and recrimination. The CASI uses extensive routing, text substitution and validation to capture data on sexual behaviour and lifestyles. Clinic staff do not have access to the patient's responses and no personally identifiable data is sought. At the outset of the study patients were interviewed about the recruitment process and experience of the CASI to allow improvements in design and to provide guidance for the interpretation of the results.

The interviews with patients used cognitive interviewing techniques to explore the process of navigating and completing the CASI. This paper discusses the efficacy of using CASI in a clinic based setting to collect, sensitive, detailed and complex data from a specific, yet diverse population group.

Keywords

CASI, sexual behaviour surveys, cognitive interviews

1. Introduction

INSIGHT (Investigation of New Seroconversions In Gay men who HIV Test) is a three-year collaborative research project funded by the Medical Research Council to investigate the current reasons why men who have sex with men (MSM) acquire HIV infection. Around 1500 HIV infections acquired through sex between men are diagnosed in the UK each year[1]. Many of these diagnoses represent longstanding infections but a substantial number are reported with evidence that infection has been recently acquired. In 2000, 13% were reported with a previous HIV negative test within the past two years. Whilst the determinants of sexual HIV exposure amongst the wider population of MSM is the subject of ongoing investigations, there exists a pressing need to identify the factors that actually result in HIV seroconversion. INSIGHT aims to do this, using subjects recruited from Sexual Health clinics, where the majority of recent HIV infections in MSM in the UK are diagnosed. Three related studies are being conducted: a case control study to identify risk factors for HIV seroconversion; a qualitative interview study of a sub-set of patients from the case control study; and a national descriptive survey to assess the distribution of risk factors investigated in the other two studies. Recruitment to the case control and qualitative studies began at the end of 2002 and will continue for two years.

2. Methods

Case Control Study

This study aims to recruit 300 MSM over a two-year period that have recently tested for HIV having had a previous negative HIV test within the last two years. It compares men who have recently tested HIV positive (cases) with men who have remained HIV negative (controls) to identify risk factors for seroconversion. The basic design had been piloted at a central London Sexual Health clinic in the mid 1990s[2]. In this study, subjects were interviewed 'face-to face' using paper questionnaires, with self-completion for the more sensitive topics. Although the methodology proved successful, the number of patients recruited was too small to fully investigate the range of factors intended. To recruit an adequate sample size, a multi centre study was required. Surveillance data indicated that the majority of HIV diagnoses in MSM were made at a handful of Sexual Health clinics located in cities with large populations of gay and bisexual men. Seven of these clinics in Brighton, Manchester and London agreed to participate in the INSIGHT case control study.

The questionnaire was developed over a six-month period through collaboration with staff at CDSC and researchers with considerable expertise in gay men's sexual behaviour questionnaires. Key topics of interest were identified and specific questions to address these developed. Questions were selected according to the following guidelines:

- The same questionnaire for cases and controls
- Appropriate and acceptable to all subjects
- Allow comparisons with external survey data where possible
- Capable of capturing the range and diversity of behaviours and lifestyles
- Quantitative
- Anonymous
- Keep questionnaire length to a minimum

The majority of questions relate to the time between the subjects last two HIV tests, during which the cases acquired HIV infection and the controls did not. The topics investigated include changeable demographics, detailed sexual behaviour, acquisition of new sexual partners, other sexually transmitted infections, general health and well being and attitudinal scales relating to safer sex and optimism around therapeutic advances for the treatment of HIV infection.

Computer Assisted Self Interview (CASI) was chosen as a means of data collection for a number of reasons. It had been successfully used to collect sexual behaviour and lifestyle information as a component of the second National Survey of Sexual Attitudes and Lifestyles[3] and in a study of HIV positive homosexual men attending an outpatient clinic in London[4].

CASI's ability to interact and validate responses meant an interview tool could be developed that shared many of the advantages of a 'face to face' interview (e.g. interaction and validation) whilst preserving the subject's anonymity, which was likely to be an important feature for the men the study aimed to recruit. CASI would also standardise interviewer bias and reduce costs by not having to employ trained interviewers at each clinic. Another persuasive factor was the fact that gay and bisexual men were increasingly turning to the Internet as a means of communication and socialising, and therefore many would be familiar with this format.

SNAP (version 6) software was chosen to construct the CASI because it provides a simple interface through which an HTML program could be developed without detailed knowledge of programming. The design followed many of the recommendations presented at the ASC Challenge of the Internet Conference in 2001 concerning the design of lengthy Internet questionnaires[5].

Key features include:

- Simple presentation
- Drop down menus
- Automatic routing
- Text substitution
- Validation
- Progress bar and navigation buttons

Despite a very long (204 main questions) and complex questionnaire, completion times of around forty minutes were achieved through the use of extensive routing and drop down menus. Text substitution was used periodically to remind subjects of the dates of their last two HIV tests. Apart from a few key questions, the option to require subjects to enter a response to every question was avoided, since this is likely to cause annoyance and generate erroneous data. Similarly the use of internal validation checks was kept to a minimum. Although it was likely that many of the respondents would be familiar with using computers, an optional mouse tutorial was included at the start of the questionnaire.

An initial IT assessment of the clinics revealed that they all required additional computers in order to participate. Desktop computers were the preferred option since compared to laptops they are cheaper, easier to operate and less likely to be stolen. However most clinics required the latter since they could not provide a designated room in order to conduct the CASI. The HTML files for the CASI operate from the hard drives of the computers, and access to the Internet (in order to submit responses) provided by dial-up or networked connections.

3. Progress

Recruitment

Recruitment to the case control study has been going well, but has not been at the levels anticipated at all clinics. In order to maximise recruitment, clinic staff were invited to attend a Collaborators Day in March 2003 to compare different approaches. Staff agreed that patient's responses to the study had been good and that low recruitment rates were due to patient's not being approached to take part in the study because:

- Problems maintaining momentum and remembering to recruit
- Staff finding it difficult to make time for recruitment
- Limited space for interviews within clinics
- Lacking confidence to approach patients
- Different studies competing for similar patients

Having identified these barriers, participants developed innovative approaches to dealing with the problems they had been experiencing. Collaborators felt that recruitment would not always be easy, but they valued the INSIGHT study and felt the results would be worth the effort required.

Lost data

Despite the many merits of the INSIGHT CASI, the potential for electronic data to be lost in transmission is a considerable drawback. Periodically data submitted through the INSIGHT CASI has not been received at CDSC. Although rare, such instances have caused great worry to the researchers and badly affected the moral of clinic staff and their confidence in the study, particularly since they often work late in order to find time to oversee the CASI. Prompt investigation of instances of lost data has revealed the following problems:

- Blocking email with 'undesirable words'
- Unreliable email service
- Failed dial-up connections

The CASI contains a few free text fields in which subjects can write what they like. Given that this is a sexual behaviour survey, subjects often express themselves in terms that might be considered 'undesirable' by corporate email scanning systems. For INSIGHT this problem was overcome by negotiating an email address that was exempt from this scanning. The problem of an unreliable email service was partially solved by routing data to an independent web mail address that has proved to be more reliable, and automatically forwards messages to CDSC. Failed dial-up connections have mostly been resolved by persuading clinics to provide networked connections where possible. In the longer run, hosting the CASI on a Website rather than being held on computers in the clinics will hopefully minimise data losses.

Cognitive Interviews

On the basis of the data received so far, the CASI appears to be functioning well, with non-response for optional fields at less than 5%. However inconsistencies are apparent in the sexual behaviour reported by some respondents. The opportunity to follow up subjects that have completed the CASI for the qualitative study has been used to conduct cognitive interviews amongst the initial recruits to the case control study. These focused on the recruitment process and subject's experiences of the CASI in order to allow improvements in the design and to provide guidance for the interpretation of results. Thirteen respondents were interviewed, including 5 cases and 8 controls aged from 20 to over 60 years with a range of socio-economic and educational backgrounds and sexual lifestyles. Three respondents spoke English as a second language.

What are cognitive interviews?

Cognitive interviewing techniques are diagnostic tools used to understand the process of replying to survey questions from the respondent's perspective. Drawing on qualitative methodology, cognitive interviews seek to encourage respondents to verbalise what goes through their minds during the process of answering questions and navigating through the CASI. These interviewing techniques explored the stages of replying to an interview question:

- Comprehension – establishing what the question means
- Recall – retrieving available information to make a reply
- Judgement – deciding what information is relevant
- Response – selecting a response

Key findings

The interviews revealed that inconsistencies in the sexual behaviour mainly arose because subjects misinterpreted the language used in the questions. It had been decided to use medical terms to describe sexual behaviours since these were considered more precise and appropriate to this type of study. However many subjects were unfamiliar with the terms *insertive* and *receptive* in relation to anal and oral sex. The use of the term *sexual partners* to describe the men that subjects had sex with led to confusion around new or casual contacts, since for many this term implied a committed relationship.

Although the CASI screen uses topic bars that provide the context for questions and definitions of the terms used, subjects tended not to read these. Expectations about what the study was about led some to select responses that they believed were appropriate, such as only reporting high-risk behaviour. It appears that most subjects are eventually able to work out the intended meaning of the questions, but this often involved considerable thought and for some, resulted in navigating back through the questionnaire to find definitions. This increased the length of time it took to complete the CASI, time which they were less likely to spend later on in the questionnaire. An updated version of the CASI is underway that incorporates these findings, using clearer language and presentation.

The main reason for subjects taking part in the study was that the clinic staff had made it sound interesting. They were keen to participate as they had good relations with the clinic staff and saw it as a way to respond to the help and support they had received. When asked about what they perceived as the benefits of the study, most reported that they saw its role in preventing the spread of HIV, and recognised that it was progress and change in the face of an evolving epidemic. Many had strong views about HIV prevention, and were critical of the early AIDS information provided in the 1980s.

Many would like to know more about the risks of sexual behaviours, particularly oral sex, and felt that it wasn't enough just to be told to use condoms all the time. In participating in this study they saw it as a means by which their own views and experiences could inform future information and interventions.

4. Conclusions

This study has demonstrated that Computer Survey Technology can be successfully introduced in to Sexual Health clinics to facilitate research. For INSIGHT, this success has mostly been achieved because the study seeks to investigate a very important topic and clinic staff and subjects perceive its value. As a result, the study has overcome the immense challenges of conducting additional research at clinics already stretched to capacity by increasing patient workloads, with limited IT experience, and subjects have been willing to spend addition time at their clinics to complete the CASI to the best of their abilities. Computer Survey Technology has facilitated this process by providing an innovative and efficient means of data collection. The use of CASI has enabled the collection of complex, personal and sensitive data as accurately as possible and with acceptable completion times. As was suspected, the majority of subjects are computer literate (less than 10% used the mouse tutorial) and over 60% have reported using the Internet.

The instances of lost data have been the biggest problem encountered in the use of this technology. Such losses may be less of a concern for large-scale Internet based studies, were they may even go unnoticed. However studies such as INSIGHT, that invest considerable resources to collect large amounts of data from relatively few subjects need to address this issue. Negotiations with the software company to tailor the CASI so that it can save data locally resulted in prohibitively expensive quotes. In the longer run it might be more cost effective to employ a programmer to design a CASI and administer it through a Website.

A key feature of the INSIGHT study is that it combines quantitative and qualitative research methods in its design. The ability to draw upon qualitative skills to conduct the cognitive interviews has brought considerable benefits, allowing improvements and fine tuning in the CASI design and guiding the interpretation of the data. We recommend that other researchers consider these techniques, particularly at the pilot phase.

References

PHLS AIDS and STD Centre - Communicable Disease Surveillance Centre, and Scottish Centre for Infection & Environmental Health. Unpublished Quarterly Surveillance Tables No. 58, 03/03 Table 6a.

Gilbart VL, Williams DI, Macdonald ND, *et al*. Social and behavioural factors associated with HIV seroconversion in homosexual men attending a central London STD clinic: a feasibility study. *AIDS Care* 2000; **12**(1): 49-58.

Fenton KA, Korovessis C, Johnson AM, *et al.* Sexual behaviour in Britain: reported sexually transmitted infections and prevalent genital Chlamydia trachomatis infection. Lancet. 2001 **358**(9296):1851-4.

Stephenson JM, Imrie J, Davis MMD, *et al.* Is use of antiretroviral therapy among homosexual men associated with increased risk of transmission of HIV infection? *Sex Transm Infect* 2003; **79**:7-10.

Wojtowicz T. Designing Lengthy Internet Questionnaires: Suggestions and Solutions. *Proceedings of the ASC Challenge of the Internet Conference* May 2001. p. 25-32.

About the Authors

Neil Macdonald is the Project Co-ordinator and a grant holder of INSIGHT. He a State Registered Clinical Scientist (Epidemiology) and has worked in HIV/AIDS surveillance for several years, primarily focusing on HIV seroconversions in men who have sex with men. He can be reached at: Behavioural Surveillance Section, HIV & STI Division, CDSC, Health Protection Agency, 61 Colindale Avenue, London, NW9 5EQ *+44 (0)20 8200 6868 ext 4546* neil.macdonald@hpa.org.uk.

Gillian Elam is the INSIGHT qualitative researcher. She has specialised in qualitative research for several years and the development of cognitive interviewing techniques, for example to test the development of interviewer administered and self-completion computer assisted interviews for the second National Survey of Sexual Attitudes and Lifestyles. She can be reached at: Behavioural Surveillance Section, HIV & STI Division, CDSC, Health Protection Agency, 61 Colindale Avenue, London, NW9 5EQ *+44 (0)20 8200 6868 ext 4548* gillian.elam@hpa.org.uk

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Mobile Phones and Telephone Survey Methods

Vesa Kuusela

Abstract

The methods applied in telephone surveys have been based on the idea that households usually have one fixed landline phone. The popularity of mobile phones has increased very rapidly especially in Western Europe and varies considerably from country to country. This has both indirect and direct effects on telephone survey methods and practice, e.g. survey costs, sampling related issues, non-response and estimation. Nevertheless, very few reports refer to the potential impacts that the mobile phones may have in surveys. This paper has two parts: First the structure of telephone coverage in Finland will be described. The change of the structure in Finland has been one of the fastest and the first part serves as an example because it is possible that similar change will take place in other countries as well. The second part deals with the influences mobile phones have had in Finland and what potential impact there may be still ahead.

5. Introduction

The methods applied in telephone surveys have been based on the idea that households usually have one fixed landline phone. For those households, which own multiple phones, there are specific handling instructions (see e.g. de Leeuw and Collins, 1997, Groves et al., 1988 or Piekarski, 1996). Especially, the methods in Random Digit Dialling (RDD) rely on the assumption that households have only fixed phones.

The popularity of mobile phones has increased very rapidly especially in Western Europe. For example in Finland, the amount of mobile phones has increased and amount of conventional phones decreased nearly ten years now. Though, the popularity of mobile phones – and conventional phones – varies very much between different population groups. In some groups, 99% have a mobile phone while in some group only 10%. A new feature is that more than one third of households have only mobile phones and the proportion of these mobile phone only households is increasing.

Popularity of the mobile phone varies considerably from country to country. For example, according to an Oftel report (2002), 68% of households in the United Kingdom (UK) had a mobile phone and the proportion had remained about the same for some time. In addition, 70% of mobile phone users in the UK have prepaid packages. Nicolaas and Lynn (2002) reported that only 5% of households in the UK had mobile phones only and 93% had fixed phones. However, most countries do not have accurate figures concerning telephone coverage, which makes it difficult to estimate how widespread problems mobile phones might introduce for survey research.

Usually the statistics describing the telephone coverage are based on data collected by telephone operators or on estimates attained as side products from surveys designed for other purposes. As a

result, the knowledge concerning the structure of telephone coverage may be fairly superficial or even inaccurate. It should be noted that telephone coverage and its structure, not to even speak about the usage of phones, could be estimated only through a survey designed for this purpose.

When the popularity of mobile phones increases in a society, it has both indirect and direct effects on telephone survey methods and practice, e.g. survey costs, sampling related issues, non-response and estimation.

Very few reports, so far, refer to the potential impacts that the mobile phones may have in surveys. It is not surprising that the bibliography made by Khurshild and Hardeo (1995) had no entries that dealt with mobile phones, but as late as in 1999, Collins made only a brief comment on the growing popularity of mobile phones in the future. Nathan (2001) and Nicolaas and Lynn (2002) recognise the problem mobile phones introduce in household surveys, especially in context of random digit dialling. Fuchs (2000) conducted an experimental survey to compare non-responses in mobile phone and fixed phone surveys.

In the U.S. where telephone interviews are very popular, the research concerning the potential influences of mobile phones on surveys has emerged (see e.g. Steeh, 2003 or Traugott and Joo, 2003).

In Finland, the structural change of telephone coverage has been followed with three similar surveys focused on this topic (in 1996, 1999, and 2001). In addition, telephone coverage is followed on monthly bases in a consumer barometer survey. In CATI center, the phone type of the interview has been recorded for three years. The results of these surveys and the influence of mobile phones in the survey quality have been reported by Kuusela and Notkola (1999), Kuusela and Vikki (1999) and Kuusela and Simpanen (2002).

This paper has two parts: First the structure of telephone coverage in Finland will be described. The change of the structure in Finland has been one of the fastest and the first part serves as an example because it is possible that similar change will take place in other countries as well. The second part deals with the influences mobile phones have had in Finland and what potential impact there may be still ahead.

6. Telephone coverage and its structure in Finland

The rapid change of telephone coverage in Finland started when the 2nd generation mobile phones were introduced in the mid 1990s. New models were considerably smaller and their batteries lasted longer than in the earlier models. The change of the telephone coverage started already in the first half of the 1990s, but accelerated considerably in the second half of the decade. Fixed phone coverage has been decreasing since 1993. Actually, the popularity of the mobile phone started to increase very rapidly two or three years later (see figure 1).

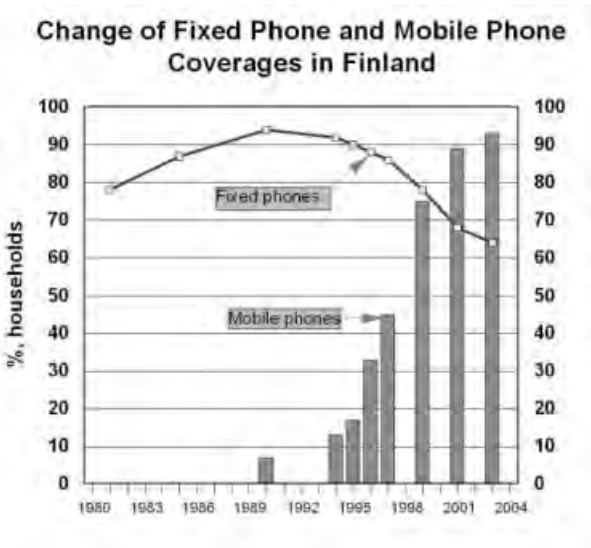


Figure 5. The trends of mobile phone and fixed landline phone coverage in Finland from 1980 to 2003

Already in 1999 the popularity of mobile phones surpassed that of fixed phones. In 2003, 93% of households had one or more mobile phones and 64% had a fixed phone. Roughly, 57% of households had both phones. A little over 25% of households had a 3rd generation mobile phone with an Internet access and email facility. A mobile phone capable of taking still and moving pictures and sending them was in 7 % of households.

A mobile phone is a personal appliance, which is carried nearly all the time, whereas a fixed phone is for the whole household and it is kept in one place. A mobile phone could be compared to a wristwatch and fixed phone to a clock on the wall. Consequently, the mobile phone coverage should be given per person rather than per household. In Finland, practically all working-age people have a mobile phone (99%), as well as younger people. Elderly people, especially elderly women, have few mobile phones (10%). In the greatest part of households, however, every member has a mobile phone of his or her own.

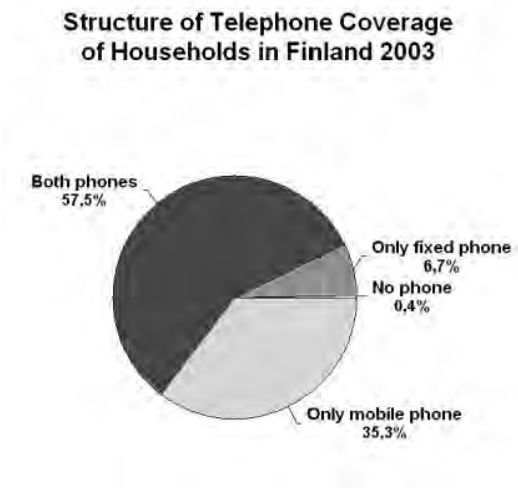


Figure 6. The structure of telephone coverage of households in Finland, February 2003

Nearly 90% of the mobile phones in Finland are privately owned and almost all mobile phones are subscribed to under monthly contracts. The use of pre-paid packages is very rare (less than 2%). In 1999, almost 90% of the mobile phones were listed either to their actual users or to someone else in the same household and that has not changed since.

The structural change has slowed down but it has not stopped, although there are signs of saturation. There is, of course, an absolute limit in mobile phone coverage but it is difficult to predict how low the fixed phone coverage will go. Typical feature is that more and more households are without fixed phones. In February 2003, more than 35% of households had only mobile phone(s), two years earlier 30%, and in 1999, 21% did not have any fixed phone. An inverse trend can be seen in the proportion of households who have only a fixed phone (7%, 12% and 23%, correspondingly)

Three different processes are influencing the increase of mobile phone only households. Some households simply give up their fixed phone and buy a mobile one, some households moving from one dwelling to another no longer get a fixed phone, and many newly established households do not get a fixed phone at all because they already have one or more mobile phones. A fixed phone is becoming a kind of luxury for many households, unless it is needed for some other purpose, like Internet access. That is, a mobile phone is found more important than a fixed phone, or the fixed phone is found superfluous.

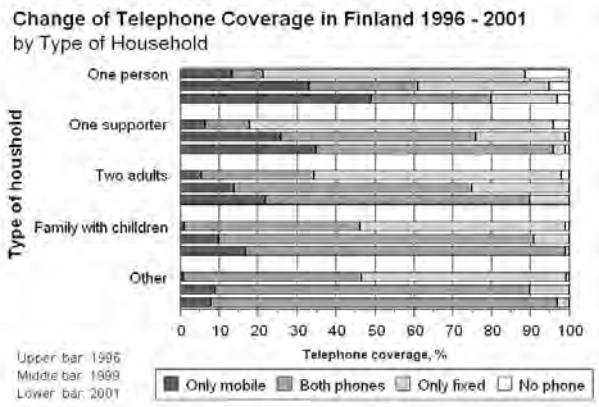


Figure 3. Change of telephone coverage from 1996 to 2001 by type of household. The "other households" are similar to families but have more than two adults

Different population groups show very different patterns in acquiring mobile phones and giving up fixed ones. In figure 3 is shown how the telephone coverage has changed in different household types.

The most dramatic change has taken place in one-person households: while in 1996, 70% of them had a fixed phone only, in 2001, 80% had a mobile phone and nearly 50% had a mobile phone only. Mobile phone coverage has also increased considerably among other types of households, but they have not given up their fixed phones to the same extent as one-person households have.

In 2001 more than one third of the households in Finland were one-person, or single adult, households, and their proportion has been growing slowly. There are two very different types of one person households: young adults who have moved from their parents' home e.g. for studies; and older widows who have not remarried. The lifestyles and living conditions of these two types vary considerably, and they are different from other types of households as well.

Figure 4 shows how telephone coverage of single adult households changed by age and gender. Roughly 90% of the youngest people, both males and females, living alone have only mobile phones and none have only a fixed phone. The proportion of mobile phone only households decreases steadily with age, faster with women than with men. In the oldest age groups, fixed phone only households are more frequent. Especially older females living alone rarely have a mobile phone and only very few have only a mobile phone.

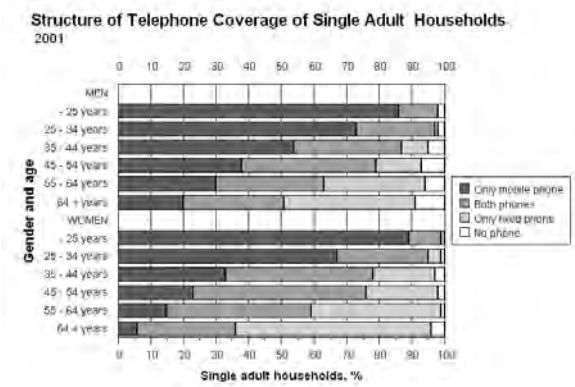


Figure 4. Structure of telephone coverage in single adult households in Finland, by gender and age

The age structure of a household is connected with the types of telephones it has. Use of the mobile phone is characteristic of young people while few older people have mobile phones. For example, in households where all members were aged less than 35 years of age, 70% had a mobile phone only and 1% a fixed phone only. The proportion of households with mobile phones only decreases as the age of the oldest member of household goes up. The frequency of fixed phones increases in households with older members. Older people have mobile phones, too, but they are not their only means of communication.

The stability of the life situation of a household is probably the most determining factor of telephone ownership. Nearly 80% of the households owning their dwelling have a fixed phone irrespective of the type of the dwelling. Mobile phone only households are relatively rare in owner-occupied dwellings. On the other hand, more than half of the households living in rented dwellings have only mobile phones.

It should be borne in mind, however, that these characteristics are not isolated but depend on each other. For example, young people may not yet have permanent dwellings but older people and families have probably lived in the same dwelling for a long time and therefore they have a fixed phone which they have not given up. The type of the building where a dwelling is located is, naturally, strongly connected both with the wealth of a household and its life situation. For example, households living in single-family houses are usually fairly wealthy. On the other hand, greater variety of households lives in blocks of flats.

The proportion of mobile phone households increases as the total household income gets higher but the proportion of mobile phone only households is much higher among low-income households. That is, people with lower incomes tend to have only mobile phones (rather than fixed ones), whereas affluent people tend to have both. Recently, however, there has also been a slight tendency among affluent households to give up fixed phones. In addition to the housing conditions, the region where a household is living influences the popularity of the fixed and mobile phone. In general, the mobile phone is more popular in cities whereas most households in rural areas have a fixed phone.

7. Influence of mobile phones on surveys

Telephone survey methods and practices have been designed for fixed phones alone (see e.g. Groves et al., 1988, Piekarski, 1996). The increasing popularity of the mobile phone requires making alterations in both. The scope of the required changes depends basically on the local conditions. There are great differences from country to country in the infrastructures of mobile phone installations and tariffs. Therefore, the usages of mobile phones also differ, as well, and consequently the needed changes vary. Nevertheless, no survey organisation in any EU country could disregard mobile phones, anymore.

Mobile phones have both indirect and direct effect on telephone survey methods and practice. Those issues, which need to be paid attention at, may be roughly divided in sampling related and non-sampling issues.

Non-sampling effects

Survey costs

In Finland, one of the most prominent consequences of the increasing popularity of mobile phones has been that also the proportion of mobile phone interviews has increased steadily. Already in July 2001, more than 50% of CATI interviews were conducted over a mobile phone and upward trend has been continued ever since. The proportion of mobile phone interviews has slight seasonal variation, though. Typically, it is highest during popular holiday seasons and lower at other times.

A call to a mobile phone is more expensive than a call to a fixed phone. Therefore, an increase in the number of mobile phone interviews increases the telephone expenses of a survey. For example, at Statistics Finland the telephone costs of interviewing rose more than 50% in the four years from 1999 to 2002.

Survey practice

All interviewers of Statistics Finland are instructed to always try to call the fixed phone when a respondent is known to have one. However, respondents are often easier to reach by the mobile phone even when they have a fixed phone. Therefore, interviewers need special instructions for mobile phone interviewing. For example, when a call is made to a mobile phone, the interviewer should always ask first whether the respondent is in a place where he or she can speak freely and ask whether an appointment could be agreed.

Even if a respondent agrees to be interviewed, the interviewer should suggest an appointment if the respondent is abroad, because half of the cost of the call will be borne by the respondent, or if the respondent is driving a car or is in a noisy environment or it can otherwise be assumed that the respondent would not be able to concentrate properly.

Survey results

Analysing whether mobile phones influence survey results is very difficult. The population groups using primarily the mobile phone or the fixed phone are so segregated that observed differences may either be caused by the type of phone or by the differences between the population groups. However, no reason has been found to support the assumption that the results would be different because of the different technical interviewing devices. This observation has been made in two independent studies (see Kuusela and Notkola, 1999 or Kuusela and Simpanen, 2002).

Apart from whether the phone type as such has an impact on the results, the type of the phone induces indirect effects that originate from the structural differences in the telephone coverage. The reason in most cases is that households in a more stable situation are more often interviewed over the fixed phone because they more frequently have those phones and they are more often at home (see Kuusela and Simpanen, 2002). Traugott and Joo (2003) observed that the political attitudes of mobile and fixed phone users are different. They concluded that is was caused by the strong relationshipt between social economic status and mobile phone ownership.

Length of interviews

A few years ago, when mobile phone devices were technically less developed than they are today, it was suspected that mobile phones had shortened the interviews. The batteries of the first models did not last long and, therefore, respondents might have answered hastily. The fact that respondents could be virtually anywhere was also suspected of having an influence on the lengths of interviews.

However, even the first analysis in 1999 failed to reveal any observable differences in the lengths of interviews between phone types (see Kuusela and Notkola, 1999). Later results have confirmed that no consistent differences can be observed, not at least in short interviews, lasting only a few minutes (see Kuusela and Simpanen, 2002). The real effects may still surface in a longer interview. Steeh observed that in the U.S. the length of interview was a little longer in mobile phone interviews than in RDD interviews (Steeh, 2003).

The length of interviews is also a difficult topic for research because respondents' different life situations may mean that they are asked very different sets of questions.

Non-response

When the popularity of the mobile phone started to increase, it was also feared that respondents would refuse from an interview over a mobile phone more frequently than from one over a fixed phone. There is no indication of this, however. Although the total refusal rate seems to remain practically unchanged, there is faint evidence of interaction: when contacting young respondents by mobile phone the refusal rate tends to be lower than that obtained with fixed phone. The situation is reversed when contacting older respondents.

Apart from refusals the popularity of mobile phones presented a different consequence, as well. Those respondents who previously were difficult to reach (e.g. young men, unemployed people, students, etc.) now have mobile phones and are reachable. That is, mobile phones also decrease non-response. The high non-response rates obtained during the holiday season have also improved considerably thanks to the mobile phone.

A similar phenomenon could probably also be observed in catalogue frame surveys[1], although it is not as easy to demonstrate. Fuchs (2000) conducted an experimental study to compare non-response rates in mobile phone and fixed phone interviews. He observed that the refusal rate was slightly lower in mobile phone interviews than in fixed phone interviews but more call attempts were necessary to reach an adequate overall response rate. Also, Steeh observed that in the U.S. the refusal rate was slightly lower in mobile phone interviews than in RDD interviews (Steeh, 2003).

[1] Catalogue frame survey means here a survey where the sampling frame is composed of a explicit or implicit (like in RDD) list of telephone numbers. Primary sampling unit is this case is a telephone number.

Sampling related effects

The sampling related problems vary between different sampling methods. If the sampling units are something else than telephone numbers, e.g. persons or households, the main problem is finding a telephone number (or numbers) for each unit. In this case, the pivotal question is how comprehensively mobile phones are listed and whether they are listed to the persons who are actually using them. In general, in such population or household samples mobile phones probably assist the survey work if mobile phones are listed comprehensively. That is, mobile phones facilitate telephone surveys and might decrease non-response because some parts of the population, which earlier were difficult to reach, can now be interviewed by a mobile phone.

If the sample frame is a list of telephone numbers or numbers are generated, like in the RDD method, both sampling and estimation may become much more complicated than before. In these so-called catalogue frame surveys sampling problems may be difficult and new methodological research will be needed to solve them.

Frame coverage

The parts of population with mobile phones only and fixed phones only are segregated: roughly speaking young people in a more or less unstable life situation usually have mobile phones only, while elderly people having a stable life situation have fixed phones only, and middle-aged people with families very often have both.

An example of the situation is that in Finland the proportion of interviews completed over the mobile phone varied considerably depending on the age of the respondent. Approximately 83% of the interviews of respondents aged between 20 and 29 but only 23% of those of respondents older that 64 were done over the mobile phone. Respondents' educational level also has an influence on this. For example, in the first half of 2002, 55.8% of interviews with respondents with basic level education, 48.9% of those with respondents with middle-level education, and 40.2% of those with respondents with high-level education were done over the mobile phone.

If either of the phone types is left out of the frame, serious undercoverage will be introduced. Undercoverage will be a source of bias in all telephone surveys regardless of the applied sampling method. On the other hand, if both phone types are included, overcoverage is unavoidable. Overcoverage only hampers catalogue frame surveys.

A mobile phone is usually a personal and private appliance whereas a fixed phone is for the entire household. Therefore, broadly speaking, sampling from a mobile phone number frame makes a population sample while sampling from a fixed phone number frame results in a household sample. The problem is that the frames overlap for certain parts of the population while for some others they do not, and the frames are different in nature. Therefore, the conventional dual frame approaches are not valid.

In some telephone surveys, all members of a household should be interviewed. Mobile phones make this design very difficult because the members of a household may be dispersed. Also designs where the respondent is selected randomly amongst the members of the household will become difficult to put into practice.

There are practically no means (in a survey organisation) to find out the whereabouts of a mobile phone. The number usually does not indicate where the user of the phone lives or works, and even if it did the phone could still be anywhere. Accordingly, all spatial sampling designs are unfeasible.

If the number space is homogenous in the sense that it is not possible to distinguish the phone type from the number itself, estimation and weighting get even more difficult.

Inclusion probabilities

In catalogue frame surveys, mobile phones will change the inclusion probabilities of households in both population and household surveys. In large households, there are presumably more mobile phones, in addition to a fixed phone, than in smaller households. The inclusion probabilities will increase linearly with the number of phones in the household. Consequently, large households will be over represented in samples.

The different inclusion probabilities maybe are not a big problem in population samples if population statistics concerning household sizes are available. However, significant intrastrata correlation may cause bias. In household samples, the varying inclusion probabilities may cause serious bias and rectifying it calls for special weighting schemes. The construction of the weighting scheme is not straightforward because of the different statuses of the phones.

Profiles of mobile phone users

Mobile phones do not present as homogenous a group as fixed phones as to their ownership and usage. In Finland, for example, although most of the mobile phones are privately owned, there are also mobile phones provided by employers to their employees who may use them for private purposes as well. They may be listed to the person or to the company. However, many people have both a mobile phone of their own and a company mobile phone.

The owners of private mobile phones are often not their users. For example, the owner might be the head of the household but the user is a child. The phone may be listed either to the owner or to the actual user. Many mobile phone users are young, some very young (under 10 years). Therefore, it has to be confirmed that the respondent is eligible if it is not known who he or she is.

So-called pre-paid mobile subscriber packages are popular in many countries. This means that the phone number is only valid for the period the prepaid fee covers. In some cases, it is possible to pay to keep the number operational but in others the number is only valid for as long a period as the paid fee covers.

8. Discussion

Mobile phones are an inherent part of infrastructure in most developed countries and some underdeveloped countries may leap over the landline technology and assume mobile phones. In some countries the change has been very rapid while in some others, such as the United States, it has been slower mainly because of tariff policy. Mobile phones are becoming more popular in the USA, albeit not very rapidly. In USA mobile phones are usually not included in RDD samples as yet (see e.g. Jenkins, 2001) but that may change in the near future. Obviously, RDD samples do not include mobile phones elsewhere either (see e.g. Nathan, 2001 or Nicolaas and Lynn, 2002).

In most European countries, mobile phones are probably already so popular that they should be taken into account in all telephone surveys. The structures of the telephone coverage and the installations and uses of mobile phone systems differ from each other in most countries and therefore a uniform methodology maybe is not feasible yet. To gain adequate knowledge raises the need for special surveys in which these issues are elucidated.

The change of the structure of the telephone coverage has some unavoidable consequences to telephone surveys. For example, when mobile phones become more popular, the number of mobile phone interviews grows automatically. An immediate consequence is that the telephone costs of surveys go up.

The future of telecommunications might be very different from what we know now. In Finland, as well as in many other countries, it is possible that speech communication will move mainly to mobile phones in the future while fixed phone lines will be used mostly for other purposes, like Internet and email. There are many reasons why this scenario is plausible in general, and young people in Finland already consider the mobile phone as their principal means of voice communication. Obviously, they won't change their habits when they grow older.

One of the most important factors in favour of mobile phones is that mobile communication technology develops very fast, whereas there has been practically no development in fixed phones for many years. Mobile phones already have a lot of functions that will not be available for fixed phones, and many new – and imaginative – functions will become available in the future. Also small, practical things, e.g. cheaper calls between two mobile phones than from a mobile to a fixed phone, also put pressure on this development. Finally, when the critical mass in mobile phone usage is reached, it will be more or less futile to use fixed phones.

Mobile phones are in many ways different from fixed phones. Typically, a mobile phone is a personal appliance, which is carried nearly all the time, whereas a fixed phone is for the whole household and it is kept in one place. Therefore, a great increase in the popularity of mobile phones changes essentially the settings of telephone surveys. In addition, the nature and topology of the number space changes and therefore, the principles of RDD have to be redesigned. For example, in the U.S. in the future it may be impossible to tell by the number is the phone a fixed or a mobile one (see Steeh, 2003). In that case the mobile phones have to paid attention to whether survey designer wants it or not.

It may happen that even greater changes in the survey methodology are ahead when the next generation mobile phones become popular. They may have features that have a greater impact than today's mobile phones on surveys. With third generation mobile phones, the line between a mobile phone and a palmtop computer will be very thin. It is possible that, in addition to other new features, they will have instant Internet and email access connected to enhanced telephone features.

When the 3rd generation mobile phones become popular, survey practices may be confronted with much bigger challenges than those presented by today's' mobile phones. In addition, new possibilities for survey research will appear. Already now in Nordic Countries there are few small-scale surveys, which are based on SMS. When mobile phones will have Internet like access, this method to carry out a survey will be an inviting option. This survey type calls for methods, which are a combination of Internet surveys and telephone surveys.

References

Collins M (1999). Editorial: Sampling for UK Telephone Surveys. Journal of the Royal Statistical Society, A, 162, Part 1, pp. 1-4

Fuchs, M. (2000): Non-Response in a Cellular Phone Survey - Experiences from a Comparison with a regular CATI Study. International Sociological Association, Research Committee 33: Methodology (RC 33), Fifth International Conference, Cologne, Germany.

Groves R. M., Biemer P. P., Lyberg L. E., Massey J. T., Nicholls W. L. Waksberg J. (Eds. 1988). Telephone Survey Methodology. New York: Wiley.

Jenkins V. (2001): The Impact of Mobile Phones on Sampling. The Frame, September 2001. Survey Sampling Inc. (http://www.worldopinion.com/the_frame/2001/sept_1.htm)

Khurshid A, Hardeo S (1995). A Bibliography on Telephone Survey Methodology. Journal of Official Statistics, Vol. 11, No. 3, pp. 325-367

Kuusela, V., Notkola, V. (1999): Survey Quality and Mobile Phones. International Conference on Survey Nonresponse, Portland, USA.

Kuusela, V., Vikki, K. (1999): Change of Telephone Coverage due to Mobile Phones. International Conference on Survey Nonresponse, Portland, USA.

Kuusela, V., Simpanen, M. (2002): Effects of Mobile Phones on Telephone Survey Practices and Results. ICIS 2002 conference papers (CD-ROM), Copenhagen.

De Leeuw, E. D. & Collins, M. (1997) 'Data Collection Method and Data Quality: An Overview' in L. E. Lyberg, P. Biemer, M. Collins, C. Dippo, N. Schwarz, & D. Trewin (In) Survey Measurement and Process Quality. New York: Wiley.

Nathan, G. (2001): Telesurvey Methodologies for Household Surveys – A Review and Some Thoughts for the Future. Survey Methodology

Nicolaas, G., Lynn, P. (2002): Random-digit dialing in the UK: viability revisited. J.R.Statist.Soc. A, 165, Part 2, 297-316.

Oftel (2002): Consumers' use of mobile telephony; Q9 May 2002. (http://www.oftel.gov.uk/publications/research/2002/q9mob_r0702.htm)

Piekarski L. (1996): A Brief History of Telephone Sampling. (http://www.worldopinion.com/reference.taf?f=refi&id=1252)

Steeh C. (2003): Survey Using Cellular Phones: A Feasibility Study. Paper presented at the AAPOR 2003 Annual Meeting.

Traugott M.W., Joo, Sung-Hee (2003): Differences in the Political Attitudes and Behavior of Cellular and Landline Telephone Users. Paper presented at the AAPOR 2003 Annual Meeting.

About the Author

Vesa Kuusela works for Statistics Finland and can be contact at vesa.kuusela@stat.fi, Box 5A, SF-00022 Statistics Finland, Finland.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

The Gold Standard Tarnishes. Is Online Research the Solution?

Dan Coates

Abstract

Participation in telephone-based research has been plummeting in North America and This negative trend is rapidly being exported worldwide. The rush toward online research has begun in North America, fueled mostly by speed and cost advantages. Additionally, online research garners better participation as a result of the respondent being able to pick their time and place of participation. However, there are signs that the greedy locusts that have reduced participation by asking too much and offering too little are feasting on this new source of participation. This paper discusses these trends and describes a primary research initiative set up by SPSS to better understand and quantify the potential impact of participation on research findings

1. Purpose

Participation in telephone-based research has been plummeting in North America, with most households erecting a fortress around the home telephone due to unsolicited contact from telemarketers as well as criminal misuse of the survey research process. Non-contact rates are past 25% … refusal rates are nearly 60% and terminations are hitting 5%. With 100% qualification assumed, only 7 out of 100 households will yield a complete using the telephone methodology. Additionally, we are finding that the scant few who participate in online research are engaged on a multitude of levels … leading to estimates that 4% of the population account for almost half of all survey research participation. Within this context, can any telephone research be considered representative?

This negative trend is rapidly being exported worldwide. Within the UK, the rise of telemarketing has created the beginnings of the same trend toward blocking unsolicited telephone contact. Once the tipping point is reached in any methodology, the ability to regain the trust of the participant is unachievable.

The rush toward online research has begun in North America, fueled mostly by speed and cost advantages. Additionally, online research garners better participation as a result of the respondent being able to pick their time and place of participation. However, there are signs that the greedy locusts that have reduced participation by asking too much and offering too little are feasting on this new source of participation. Unless steps are taken to learn from what went wrong with the last methodology, we will soon find online data collection as arid and infertile as telephone data collection.

With major capital allocation decisions being made from the results of telephone research, and more of these types of projects are being conducted online, the time has come to better understand the impact

of sampling and data collection mode preference on the general consumer's interest in survey research participation.

2. Academic Context: Prior Matter

In evaluating online research, two biases are most often cited: self select bias and coverage bias (the percent of the population that cannot be reached using a particular methodology). However, telephone research is subject to biases that have grown since it became the 'Gold Standard' of the U.S. research market in the late 1960's. Let us review these biases and their pronounced growth in recent years:

3. Contact Bias:

There is an increasing percentage of the U.S. population that can not be reached via telephone.

> *"In telephone research, the first attempt to contact the respondents can be expected to yield only 25 or 30% of the original sample, with roughly equivalent results on the second call (first callback). Beginning with the third attempt to reach respondents, the yield of completed interviews declines rapidly".*[1]

In reporting research results, most non-response bias is ignored in practice. Households that are not contactable are often not used to calculate 'participation rates'.

> *"The findings indicate that in many consumer telephone surveys a large percentage of designated respondents or households are not contacted. Further, if contact is made, the chance that a refusal will result is better than one in four. These two factors together have a dramatic effect on the response rate."* [2]

> *"Non-response can seriously bias survey estimates and distort inferences. The relationship between non-response and survey estimates is simple and well defined, but the actual impact is ill known."* [3]

The percentage of Americans without web access has fallen to 38% in the last six months according to Nielsen/NetRatings estimates. With the rise in non-contact rates as a result of call blocking technologies, it will soon be the case that Internet and telephone coverage biases are equivalent.

4. Refusal Bias:

Within the context of a busy, time starved American consumer, the notion of answering a phone call from an unknown third party for little or no economic reward is antiquated and anachronistic.

> *"Both suppliers and their clients recognize that "refusals" not only add to data-collection costs but also that they can bias sample results . . . both responders and non-responders regard telephone surveys as invasions of their privacy that take longer than they are led initially to believe. Regarding what they like about participating in telephone surveys, it is noteworthy that "nothing" was a popular response for both responders and non-responders."* [4]

---

[1] Dunkelberg/Day. "Non-response Bias and Callbacks in Sample Surveys." Journal of Marketing Research -- May, 1973

[2] Wiseman/McDonald. "Non-contact and Refusal Rates in Consumer Telephone Surveys." J. of Marketing Research -- Nov. 1979

[3] Smith. "The Hidden 25 Percent: An Analysis of Non-response on the 1980 General Social Survey." Public Opinion Quarterly -- 1983

[4] Struebbe/Kernan/Grogan. "The Refusal Problem in Telephone Surveys." Journal of Advertising Research -- June/July 1986

> *"Americans have a limited reservoir of goodwill to expend on intrusive and unsolicited telephone contacts of any kind--reasonable or unreasonable."*[1]
>
> *"Americans have a good opinion of research, but many still won't participate in phone surveys."*[2]
>
> *"Neither sample stratification nor ascription techniques mitigate the impact of low response rates on data quality. To the extent that a survey does not succeed in interviewing these active/mobile individuals, a medium's/media vehicle's audience can be significantly under or overstated."*[3]

While Internet methodologies are cited as being prone to 'self select bias' it is clear that Americans are opting out of the telephone based survey research process at increasing levels. In fact, the use of Internet technologies can reduce the 'professional respondents' that are not found within the ranks of those prone to non-contact bias and refusal bias.

> *"The important point is that we estimate much lower participation rates than previous authors who have used the same data. . . US survey participants are extraordinarily concentrated among a small sub-segment of the total adult population . . .a tiny 4% to 5% of adults account for more than half the survey responses. Thus the research industry may be burning out the small fraction of survey participants."*[4]
>
> *"Researchers can no longer ignore the possibilities that computer technology holds for the future of our profession."*[5]

5. Methodology

SPSS has conducted a primary research initiative to better understand and quantify the potential impact of participation on research findings. A multi-cell research design was employed to isolate the impact of data collection mode by simultaneously controlling for other potential influences.

The only individual who can draw a truly random sample is the warden of a prison. Unfortunately, prisoners have no access to the Internet. A close substitute came in the form of the Indiana University / Purdue University faculty directory. IPFW is based in Fort Wayne Indiana and was kind enough to provide a full faculty listing on their web site, complete with email address, telephone number and mailing address.

The 1,311 listed faculty members were randomly separated into three cells: Web, Phone and Mail with 437 members each. Members of each of these three cells were contacted using the mode that describes their category.

The data collection operation is underway, with results collected between August 18th and September 7th. Findings will be presented at the ASC conference in Warwick, UK and will focus on any impact on key measures such as non-contact, participation and termination as a direct consequence of various data collection treatments.

---

[1] Remington. "Telemarketing and Declining Survey Response Rates." Journal of Advertising Research -- May/June 1992

[2] Tuckel/O'Neill. "Call Waiting." Marketing Research -- Spring, '95

[3] Guggenheim. "All Research is Not Created Equal!" Journal of Advertising Research -- February/March 1989

[4] Bickart, Schmittlen. "Distribution of Survey Contact and Participation in the United States" Journal of Marketing Research -- May 1999

[5] Schuldt/Totten. "Electronic Mail vs. Mail Survey Response Rates." Marketing Research -- Winter 1994

About the Author

Dan Coates is based in Dallas, Texas and is the Vice President of SPSS Online. Prior to joining SPSS MR in 2001, Dan was founder and president of Eppointments.com and a co-founder of PlanetFeedback.com, both ASP-based businesses that pushed the envelope of consumer-to-business interaction and information gathering. A pioneer of interactive technologies within the marketing research community, Dan spent over a decade developing cutting edge practices and methods on behalf of Burke Marketing, Millward Brown Interactive and the Angus Reid Group, enabling companies such as IBM, Intel, Microsoft, Sun Microsystems, Viacom and Turner Entertainment Networks to develop online marketing insight. A frequent speaker at Marketing Research industry conferences and events, Dan has delivered speeches at over 30 MR industry events. Dan's educational background includes a Bachelor of Arts degree in Sociology and a Management Systems Science degree.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Effectiveness of Progress Indicators in Web Surveys. It's What's Up Front that Counts

Frederick G. Conrad, Mick P. Couper, Roger Tourangeau, Andy Peytchev

Abstract

Web surveys often display information to respondents about how much of the questionnaire they have completed. The assumption seems to be that respondents will be more likely to complete the entire questionnaire if they know how much more effort is required. We report a web survey experiment that explores the effectiveness of such "progress indicators" when progress is calculated at three speeds: initially fast but slower later, initially slow but faster later and constant speed. The feedback presented in the early part of the questionnaire turns out to be what matters most. When initial progress was slow (and presumably discouraging) respondents broke-off at higher rates than their counterparts and when it was fast (and presumably encouraging) they were less likely to break-off on a laborious item in the middle of the questionnaire even though the rate of their progress had slowed substantially. Encouraging news at the outset led those who completed the questionnaire to rate it as more interesting and to have taken fewer minutes than did initial discouraging news.

Keywords

Web surveys, progress indicators, break-off rates, feedback

1. Introduction

A friend tries to encourage you by observing "There's light at the end of the tunnel." Her comment may help you persevere because, with the end in sight, the remainder of the task becomes more pleasant or the prospect of abandoning it more unpleasant. In either case, your friend is trying to influence how you perceive the task in order to help you complete it. Or she might try to encourage you by indicating that you are moving toward your goal sufficiently fast. The belief seems to be that the task becomes more bearable when we can tell we are making progress. Presumably this is why, on long international trips, airlines sometimes provide passengers with graphical information about the plane's position on the flight trajectory. If people compare their position before and after the movie, they should notice a difference and this difference is pleasing. This kind of thinking seems, at least implicitly, to underlie the use of progress indicators in tasks involving software, including web surveys.

Progress indicators often consist of a graphical bar that changes size in accord with the proportion of the task that is completed. They commonly appear when users copy large files (like music files) from a remote computer to their own or when they install software. In these cases, it is the computer's progress that is displayed. The user is not told why progress appears to surge and stall and can do little

to affect the rate of progress. In a sense the value of these progress indicators is that they indicate the computer is still working on the task, i.e. that it has not hung. In contrast, web surveys require the user (the respondent) to carry out a task, i.e. to answer a series of questions, over which they have considerable control. Respondents may not only decide to abandon the task, i.e. to "break-off" based on the feedback they receive but they may also modify the speed at which they complete the questionnaire by putting more or less effort into their answers. So we might expect progress indicators in web surveys to have a substantial impact on respondents' performance.

The evidence about the effectiveness of progress indicators in web surveys is limited and mixed. In one study (Couper, Traugott & Lamias, 2001), there was no difference in response rates when progress indicators were used and when they were not used. Couper et al. proposed that because the progress indicator was a graphical image (similar to a pie chart indicating percent completed), the questionnaire on which it was displayed took longer, page-by-page, to transfer to respondents' computers than did a questionnaire with no progress indicator, thus mitigating any advantage from the feedback. Crawford, Couper and Lamias (2001), controlled transfer time and actually found a lower response rate when progress indicators were used than when they were not. They observed that much of the abandonment occurred on questions requiring open-ended responses, presumably a more difficult response task than selecting from fixed choices. They report results from a follow-up study in which the problematic questions had been excised from an otherwise identical questionnaire. The respondents who were given information about their progress had completed the questionnaire at a four percent higher rate than those who were not given progress information.

Part of the explanation for these mixed results may have to do with what information is actually conveyed by the progress indicator. Crawford, et al. (2001) suggest that the progress indicator may have understated actual progress thus discouraging respondents who (correctly) believed they were further along than indicated. In particular, respondents completed almost 40 percent of the questionnaire in the first 20% of the time according to the progress indicator. In general, discouraging information, e.g., that the task will take a long time or more time than expected, may well deter respondents from completing the questionnaire. And the timing of the information may matter as well. Encouraging information, e.g., that the end is in sight, will not motivate respondents who have already abandoned the task due to discouraging preliminary information. In the current study, we explore how the quality (encouraging, discouraging) and timing (early versus late) of information displayed in progress indicators affects the completion of an on-line questionnaire.

2. Current Experiment

Respondents from a commercial panel were invited by email to answer a questionnaire administered on the web concerning a variety of "lifestyle" topics. As an incentive to complete the questionnaire, panel members qualified for entry into a sweepstakes in which they could win up to $10,000 by reaching the final screen. The topics of our survey included health and nutrition, travel, cars, and expenditures, among others. A total of 39,217 email invitations to participate in the current questionnaire were sent, in response to which 3,179 persons (8%) logged into the survey and 2,722 (7%) completed it. Thus a total of 457 persons started the survey but did not complete it, representing an overall break-off rate of 14.4%. Respondents were not given any information in the invitation letter or in the introductory pages of the questionnaire about how much time would likely be required to complete it.

Our main concern was how feedback about progress would affect the break-off rate, i.e., the number of respondents who began the questionnaire but abandoned it before completing the final screen. A textual progress indicator (e.g. "17% completed") appeared at the top of each page for half the respondents. The other half of the respondents was not given any feedback about their progress. Whether or not respondents received feedback was randomly determined when they began to answer the questionnaire. The progress indicator was designed so that the time to display a screen was unaffected by whether or not any feedback was presented. A small script was executed on the respondent's computer determining what progress, if any, to display, and this script was the same size, i.e., required the same transfer and execution time, irrespective of what it displayed.

Progress was calculated in one of three ways that affected the speed with which it accumulated (see Figure 1). In all cases, progress reflected the percentage of screens, including the current screen one, displayed up to that point. Fifty-seven screens presented a question and their display incremented the progress indicator. Ten other screens displayed did not present questions or accept answers and their presentation did not affect the progress figure. What differed was how the percent progress was derived. For one type of progress indicator (Constant speed), relevant screens were numbered from 1 to 57 and progress was based simply on the current screen number divided by 57 (presented as a percent). Thus progress increased as a linear function of screens and, therefore, at a constant rate across the questionnaire. For another type of progress indicator (Faster-to-Slower) the rate of progress decelerated across the questionnaire, accumulating quickly at first but more slowly toward the end. We produced this pattern of feedback by dividing the log of the current screen by the log of the final screen (expressed as a percent). For example, after only 9 screens respondents would pass the 50% mark (progress for screen 9 is 52%) but would need to complete[1] another 36 screens to reach the 90% mark (progress for screen 45 is 91%). Thus, the feedback is more encouraging – progress accumulates faster – in the beginning than the end. Finally, for a third group (Slower-to-Faster), the rate of progress accelerated across the questionnaire, accumulating slowly at first and more quickly toward the end of the questionnaire. We produced this pattern of feedback by dividing the inverse log of the current screen by that of the final screen. For example, to reach the 50% mark, these respondents would need to complete 60 screens (progress for screen 60 is 52%) but only another 7 screens to surpass the 90% mark (for screen 66 the amount of progress displayed is 83% and it jumps to 100% for screen 67). Thus this feedback is discouraging early on – moves slowly – and gets more encouraging toward the end of the questionnaire.

[1] By "complete" we mean advance to the next screen, which respondent accomplished by clicking a navigation button. They did not have to enter a response for a given question in order to advance.

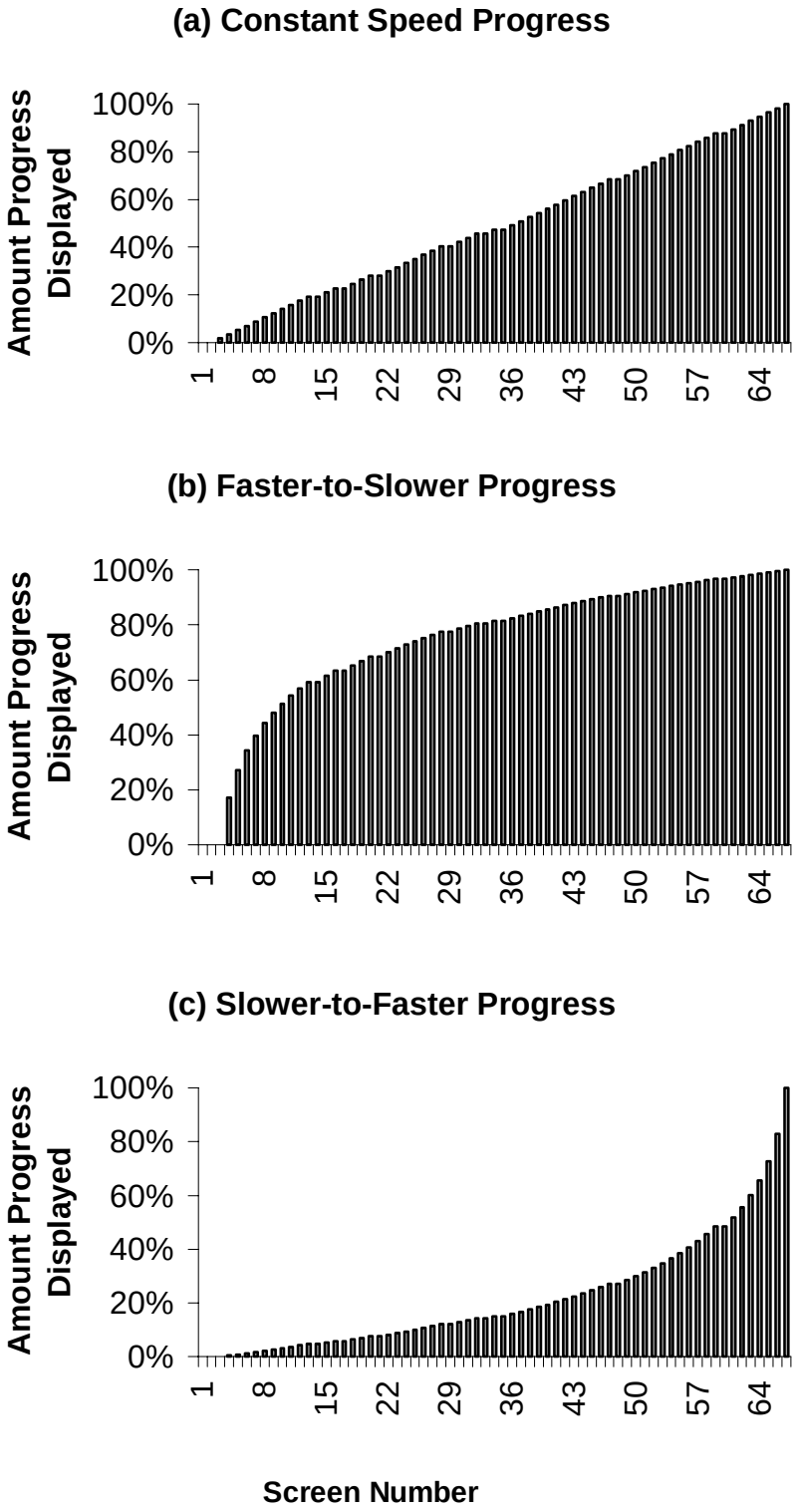


Figure 1. Rates of progress displayed in three progress indicators.

To specifically evaluate the effect of the progress indicator on completion rates, we included a particularly difficult item in the middle of the questionnaire to increase break-offs overall. Two forms of the questions were created both of which were intended to be difficult but one was intended to be easier than the other. The critical issue was whether break-off rates on the two forms differed by type of progress indicator.

The topic of the item was automobile ownership. The easier form required the respondent to register several answers by selecting radio buttons. The more difficult form required the respondent to type in several open text answers. The idea was that if the feedback in the first half had been generally positive, then there would be relatively little difference between the forms. However, if respondents had received generally negative feedback, then the difficult form would lead to substantially more break-offs than the easy form. The assignment of the form of this question was randomly determined when respondents completed the preceding screen. Thus the two main factors in the experiment were type of progress indicator (4 levels) and form of automobile ownership question (2 levels).

Questionnaire. The questionnaire was comprised of 67 screens, 57 of which presented at least one question. In general, an item consisted of a single question, but in some cases, several questions were presented in a grid format, i.e. several question stems or topics arrayed as rows and a shared a set of response options arrayed as columns. In one case – the screen about automobile ownership – multiple distinct questions appeared on a single screen. On ten screens no question was presented and these were not considered in the calculation of progress. These included nine screens on which only introductory text about the subsequent section appeared and a final screen, which indicated the conclusion of the questionnaire and thanked respondents for their participation. Respondents moved between all screens, both backward and forward, by clicking a navigation button.

Results

Break-off rates varied with the type of progress indicator, $F(3, 3176) = 10.62$, $p < .001$ (see Table 1, Row 1)[1]. Respondents were more likely to break-off when the initial feedback was discouraging (Slower-to-Faster) than when it was encouraging (Faster-to-Slower), neutral (Constant Speed) or there was no feedback at all, comparison of Slower-to-Faster to the other three progress indicator groups, $t(3173) = -5.312$, $p < .001$. Apparently, respondents receiving discouraging news at the outset reasonably assumed progress would continue to accrue slowly and inferred that the questionnaire would take more time than it actually did, i.e., more time than many were willing to invest. This could suggest that constant speed feedback for a longer questionnaire – which would resemble the initial Slower-to-Faster information for the current questionnaire– is a disincentive to continue. Even for the current, relatively short questionnaire, constant speed feedback did not motivate respondents to complete the questionnaire relative to no progress information. In fact, the proportion of respondents who abandoned the questionnaire with constant speed feedback was higher (though not significantly) than for those receiving no feedback.

Table 1.

Break-off rates, break-off screen and item non-response rates for three progress indicator groups.

Progress Indicator	None	Constant Speed	Slower-to-Faster	Faster-to-Slower
% Break-offs	12.7	14.4	21.8	11.3
Median Final Screen Number[2]	33	32	27	34

[1] An alpha level of .05 is used for all statistical tests.

[2] Respondents who broke off on the first screen (where no progress is displayed) and who completed the questionnaire are removed from this analysis.

Table 1.

Break-off rates, break-off screen and item non-response rates for three progress indicator groups.

Progress Indicator	None	Constant Speed	Slower-to-Faster	Faster-to-Slower
Mean Number Missed Items per R	1.24	1.14	1.42	1.02

Not only did discouraging news early on lead to more break-offs than did the information from the other progress indicators but there is also a suggestion that it led to earlier break-offs. We removed respondents who (1) broke off on the first screen because no progress had yet been displayed and (2) completed the questionnaire, and then analysed the final screen number before these respondents broke off. The median final screen number (Table 1, row 2) is smallest (break-offs are earliest) when discouraging information is presented first and largest (latest) when encouraging information is presented first, although a comparison of these two groups is not significant (Mann Whitney U = 19230, p = .10).

If good news up front does indeed encourage respondents to stick to the task, then they should be more likely to continue in the face of a difficult question than those who have not been so encouraged. We can test this by looking at break-offs that occur on the two forms of the automobile ownership item. The number of break-offs for this item overall was larger than for all others (aside from the first screen), though not large overall, so the difficulty of both forms increased break-offs as expected. Moreover, a larger number of respondents (.021 of all respondents) broke off on the screen presenting the more difficult form than on the screen presenting the easier form (.004 of all respondents), $F(1,2907)=11.76$, p =.001. Consistent with the idea that good news early on may lead respondents to persevere, the difference between the forms depended on the type of progress indicator, form x progress indicator interaction $F(3,2907)=2.49$, p =.058. For the Faster-to-Slower (good news first) group the difference between forms was virtually eliminated (and not reliable): there were no break-offs on the difficult form and only one on the easier form. However, the form difference was reliable for the groups that received no feedback (z=3.40, p < .001) and uniform speed feedback (z=2.62, p < .01). The fast early progress apparently continues to help respondents persevere even though the rate at which progress accumulates has slowed down substantially by the time the automobile ownership question is presented. For these respondents, completing the first screen increased progress by 18% (from 0% to 18%) but completing the 32nd screen (right before the automobile ownership screen) increased progress by only by 1% (from 80% to 81%).

It is possible that progress indicators can both motivate respondents to reach the last screen of the questionnaire, i.e. reducing break-offs, and also to be more vigilant on individual questions, i.e. reducing item non-response. Couper, Traugott & Lamias (2001) found no evidence that a (constant speed) progress indicator reduced item non-response but our variable speed progress indicators can serve as a stronger test of this proposition. In particular we might expect that, for respondents who complete the questionnaire, early positive information might lead them to skip fewer questions. As it turns out, we do not find reliable support for this proposal, $F(3,2718)=.992$, n.s. However, the average counts of skipped items are ordered sensibly (see Table 1, row 3). In particular, we observe the largest number of skipped items when discouraging news is given first (Slower-to-Faster) and the smallest number when encouraging news is given first (Faster-to-Slower). The pattern suggests to us that more items or more respondents might demonstrate a reliable advantage of early positive information for item response rates.

The effect of progress indicator type does not seem to be due to differences in how much attention respondents paid to the feedback. In a debriefing question, respondents were asked to indicate when they noticed the feedback on a 4-point scale. The scale ran from the beginning of the survey, which we have coded as 1, to the current debriefing item (screen 66 of 67), which we have coded as 4. The mean score was 1.13 suggesting that most respondents noticed the feedback very early. Moreover, there were no reliable differences between the three progress indicator groups suggesting that all respondents were equally aware of the information. Of course, only respondents who completed the questionnaire answered the debriefing questions and so we cannot say that those who broke off prior to answering this question would have registered similar responses. However, there is no *a priori* reason that those who broke off should have noticed the progress indicators any later than the complete cases, especially given the systematic effect of the different progress indicators on break-offs.

In general, respondents' self-reports measured by the debriefing questions were consistent with the picture that has emerged from the break-off data (see Table 2). First, the type of progress indicator affected how interesting respondents found the task, F (3, 2709) = 3.95, $p < .01$. Those who received good news early on judged the questionnaire to be more interesting than did those in the other progress indicator groups, comparison of Faster-to-Slower to other three progress indicator groups $t(1,2709)=125.25$, $p < .001$. Apparently people evaluate the content of the questionnaire more favourably when things initially appear to be going well than when they do not. Second, the type of progress indicator affected respondents' judgment of the length of the task, $F(3, 2709) = 4.35$, $p < .01$. The same respondents who judged the questionnaire be more interesting, i.e. those who received good news first, estimated that it took fewer minutes to complete than respondents in the other progress indicator groups, comparison of Faster-to-Slower to other three progress indicator groups, $t(1,2709)=3.12$, $p < 001$. In fact there were no differences in actual duration, $F(3,2718)= 0.59$, n.s., so respondents apparently perceived time to move more quickly when progress accumulated quickly at the outset than when it accumulated more slowly at the outset.

Third, people's judgment about the accuracy of the progress feedback is affected by the type of feedback they received, F (2, 1347) = 33.87, $p < .001$. Those who received discouraging news at first (Slower-to-Faster) reported that the progress indicator reflected their progress less accurately than those who were presented encouraging information at first ($p < .001$) and those who were given constant rate information throughout ($p < .001$). Fourth, the type of progress indicator affected respondents' judgments about how useful they found the feedback, F (2, 1352) = 9.40, $p < .001$. While none of the groups found the feedback very useful, respondents who received discouraging information first (Slower-to-Faster) reported that it was more useful than did those who received encouraging news first ($p < .001$) and constant rate information ($p < .001$)[1]. This could suggest that discouraging information is taken to heart to a greater degree than encouraging information which could be viewed as a variant on the phenomenon of "loss aversion" (Kahneman, Knetsch & Thaler, 1991), namely that losses are more painful than comparable gains. In this case, what might be lost is the satisfaction of completing the questionnaire or the opportunity to participate in the sweepstakes that offered as an incentive to participate.

[1] Note that those who did not receive any progress information did not answer debriefing questions about the accuracy and usefulness of the feedback.

Table 2.

Mean responses to debriefing questions.

Progress Indicator	None	Constant Speed	Slower-to-Faster	Faster-to-Slower
Question				
How interesting was this survey[1]?	4.09	4.03	4.07	4.27
How many minutes do you think it took you to complete the survey?	14.43	13.97	15.38	13.47
Did the progress indicator accurately reflect your progress through the survey?[2]		1.09	1.28	1.11
How useful did you find the progress indicator?[3]		2.18	2.54	2.17

Overall, the debriefing results are striking given that, by the time respondents completed these questions, the rate of progress had largely reversed for the variable speed indicators yet did not seem to reverse respondents' perceptions. It appears, from these data, that respondents form opinions about the task early on and these first impressions are not substantially modified by later evidence.

3. Discussion and Conclusion

The quality of the feedback (encouraging, discouraging, etc.) given early seems to be the critical factor in determining how these progress indicators affected respondents. If the early feedback indicates relatively rapid progress, this seems to cement respondents' commitment for the duration of the task, despite a particularly difficult question later on. Moreover, it leads them to perceive the whole experience more favorably. It is possible that the effect of the progress indicator would be even stronger with respondents who were chosen at random and generally less motivated than were ours who were self-selected and had previously agreed to participate in surveys on the Internet. It is also possible the effect of the progress indicator would have had a different form if respondents had access to other relevant information such as the likely duration of the questionnaire. Crawford, et al. (2001) report an interaction between the likely duration indicated in an invitation letter and the presence of a progress indicator. The point is that people's perception of duration and effort, their sense of boredom, and their mood, are sensitive to the information available during a task.

This had been demonstrated in several experimental studies. For example, London and Monello (1974) slowed and speeded the movement of an ordinary wall clock while subjects wrote short essays. When the clock moved more slowly than actual time, participants rated the task as more boring; when the clock moved faster than actual time they judged the task to be less boring. Lawrence, Carver & Scheier (2002) varied the accuracy scores given to participants making judgments about the meaning of artificial words where the goal was to be as accurate as possible. They indicated to some subjects

[1] 1=Not at all interesting; 6=Extremely interesting

[2] 1=Yes; 2=No

[3] 1=Not at all useful; 6=Extremely useful

that they were getting more accurate over the course of the study and to others that they were getting less accurate. Those whose accuracy appeared to increase reported that their mood improved over the course of the session and those whose accuracy appeared to decrease reported their mood changed for the worse. We should not be surprised therefore to observe different effects of progress indicators that send respondents different messages about their performance.

One implication of the current work is that, if the questionnaire is very long, a regular progress indicator may not be very effective in reducing break-offs. One could therefore make the case for presenting no progress information, although it may be deceitful to withhold progress information if it is potentially available. Alternatively, it is possible that intermittent progress feedback might reduce break-offs. Thaler (1992) found investors to be more tolerant of undulations in the value of their portfolio when the feedback was infrequent than when it was frequent. It may also be the case that verbal feedback with a motivational tone is more effective than the numerical equivalent when the latter would be discouraging, e.g. "Hang in there. You're in the final stretch."

One more consideration in designing progress indicators is whether it is fair to respondents to provide them with variable speed feedback along the lines of our Faster-to-Slower indicator. This may well increase completion rates and leave respondents feeling better about the experience than the alternatives but it may also involve a type of distortion – depending on the definition of progress. However, it could be that the subjective experience of progress is not a linear percentage of completed screens but one in which the completion of early screens is weighted more heavily than the completion of later ones. If this is so, then larger increments per screen at the outset may not distort progress at all. Moreover, it may be that respondents seek encouragement most actively at the start of the task when they are least certain about their ability to complete it. Denying them encouragement when they want it most may do them a disservice.

References

Couper, M., Traugott, M.& Lamias, M. (2001). Web survey design and administration. *Public Opinion Quarterly, 65,* 230-253.

Crawford, S.D., Couper, M. P., & Lamias, M. J. (2001) Web Surveys: Perception of burden. *Social Science Computer Review, 19,*146-162.

Kahneman, D., Knetsch, J.L. & Thaler, R.H. (1991). Anomalies: The endowment effect, loss aversion and the status quo bias. *Journal of Economic Perspectives, 5,* 263-291.

Lawrence, J.W., Carver, C. S. & Scheier. M. F. (2002). Velocity toward goal attainment in immediate experience as a determinant of affect. *Journal of Applied Social Psychology, 32,* 788-802.

London, H. & Monello, L. (1974). Cognitive manipulation of boredom. In London, H. & Nisbett, R.E. (Eds.) *Thought and Feeling: Cognitive Alteration of Feeling States.* Chicago: Aldine, pp. 74-81.

Thaler, R.H. (1992). *The Winner's Curse.* New York: Russell Sage Foundation.

About the Authors

Frederick G. Conrad, Mick P. Couper and Roger Tourangeau have appointments in the Institute for Social Research at the University of Michigan and in the Joint Program in Survey Methodology at the University of Maryland. Andy Peytchev is a doctoral candidate in the Program on Survey Methodology at the University of Michigan. This study is part of a program of research on Web

survey design funded by the National Science Foundation (Grant SES-0106222) and National Institutes of Health (Grant R01 HD041386-01A1). We thank Reg Baker, Scott Crawford, Dirk Heerwegh, Theresa Marvin, Duston Pope, and Jenefer Willem for their advice and assistance. Address correspondence to: Frederick G. Conrad, Institute for Social Research, University of Michigan, P.O. Box 1248, Ann Arbor, MI 48106, USA; e-mail: fconrad@umich.edu.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

The Effect of Images on Web Survey Responses

Mick P. Couper, Frederick G. Conrad, Roger Tourangeau

Abstract

The World Wide Web offers many opportunities for innovative design of survey and experimental instruments. One of these aspects is the ease with which rich visual images such as photographs can be used. Images may enhance motivation and increase enjoyment; they may also facilitate the respondent's task of understanding and responding to the question. However, images may also distract respondents from the task or suggest interpretations that may diverge from those intended by the survey designers. We are conducting a series of experiments to explore the possible effects of such images on the answers respondents provide in Web surveys. Specifically, we are varying both the content of the images and their placement. We are examining both subjective measures (e.g., depression) and behavioural measures (e.g., travel frequency and destinations). This paper presents the results of some of the initial studies and offers some guidance on the use of images in Web surveys.

Keywords

Web surveys; images; measurement error

1. Introduction

One of the major potential benefits of the World Wide Web as a survey data collection tool is the ease with which rich visual information can be used. This ranges from the use of colour and variations in layout and design, to the inclusion of images such as drawings or photographs in the survey instrument. While such devices have also been possible in paper-based surveys, several factors mitigated against their use. These included 1) the cost of reproducing colour images, 2) the difficulty of randomizing images or questions, and 3) the difficulty of controlling the sequences of display. These issues are trivial in surveys conducted via the Web. While images are sometimes used in computer assisted self-interviewing (CASI) applications (usually involving stand-alone laptop computers), such use has been relatively rare. The widespread adoption of the Web for survey data collection is likely to usher in a dramatic increase in the use of images in surveys.

Tempting though it may be to include images and other visual devices in Web surveys, there is much we still don't know about the effect that such images may have on the responses people give to survey questions. While there is a large and growing body of research on visual information processing in the fields of learning and communication (e.g., Lohr, 2000; Zillman, Knobloch, and Yu, 2001), this area remains largely unexplored in the survey setting. With this in mind, we are embarked on a program of research to explore how images and other visual information influence respondents in surveys. In this way, we can learn to more effectively utilize this powerful tool to enhance the quality of the data we

collect (i.e., reduce measurement error) and enhance the motivation of respondents completing the survey.

2. Visual Information

Images are one form of visual information. At its broadest, visual information is everything one sees with the eyes. But typically in the visual design field, visual information refers to all the extra-text (or non-verbal) information. It includes how the text is presented, but not the content of the text itself (the words). Horn (1998, p. 8) defines visual language as "the integration of words, images, and shapes into a single communication unit." When the words and images are congruent and convey the same basic information, they can be viewed as a single communication unit. But what about images that are part of the visual field, but are not intended to be viewed as related to the text? To quote Horn again, "[What] happens to our thinking process when we place next to one another visual and verbal elements that are not obviously related or not related at all? Almost all of us will struggle valiantly to create meaning out of nothing, to understand the close juxtaposition of the idea conveyed by the visuals and the words. … We must be vigilant in visual language, perhaps more so in prose, because readers will attempt to make meaning out of elements whose proximity may be accidental" (Horn, 1998, p. 111). Linking this caution with the context effects literature that suggests all information is considered by the respondent as potentially relevant to the task at hand (e.g., Schwarz, 1996; Schwarz and Sudman, 1992; Tourangeau and Rasinski, 1988), we argue that the use of graphic enhancements ranging from colour choices to rich visual embellishments such as colour photographs in Web surveys may affect the interpretation of the survey question or shape the response in other unintended ways. In this way, the use of visual information may produce measurement error in Web surveys.

Writing in the context of multi-media instruction, Duchastel and Waller (1979, p. 21) note that illustrations or images can be classified according to the potential effects they have on the reader (or respondent in the survey case). They identify three main roles: 1) attentional (to draw attention to the information being conveyed), 2) explicative (to add explanation to the textual material), and 3) retentional (to enhance recall of information). In similar fashion, images may affect survey respondents in different ways. In some cases, as Horn (1998, p. 225) notes, "images have the power to evoke strong, immediate feelings as well as to subtly insinuate a background climate or overall mood." Images may thus not only activate a mood or emotional state, but may also trigger attitudes or evaluations that may shape responses to associated questions (e.g., Huddy and Gunnthorsdottir, 2000; Klauer, 1988). Images may also affect the more conscious processing of the intended meaning of the survey question. For example, the presentation of certain exemplars may lead to assimilation or contrast effects. Pictures may suggest a broader or narrower category membership than intended by the question (e.g., Macrae et al., 1999). Pictures may stimulate recall of certain events that the respondent might otherwise have forgotten (e.g., Prior, 2002). Pictures may suggest certain desirable answers. In other words, the inclusion of images in Web surveys may affect all stages of the survey response process.

Thus, the effect that an image may have on the survey question may depend on a number of factors, including the type of picture (i.e., its content) and the type of question (e.g., a judgement or attitude item, a behavioural frequency measure, etc.), the degree of picture-question congruence (how the two are semantically related), the degree of visual proximity (the syntactic or structural linkage between image and question), the visual prominence of the picture (size, placement, brightness, clarity, etc.) and so on.

This brief review is not meant to be an exhaustive discussion of the role of images or visual information in Web surveys; rather, it is to point to the potential complexity of the problem. As the well-known aphorism goes, "a picture is worth a thousand (or ten thousand) words." In other words, the visual richness of the content of images further complicates the issue—we do not all perceive or interpret images the same way. The effects may thus well not be uniform or consistent across respondents. With this in mind, we present some initial efforts to understand the role that pictures may play in the answers to Web survey questions. But, as with the early question wording literature, demonstrating that effects can and do occur is only the first step—to be useful, we need to understand the mechanisms by which these effects operate, and the conditions under which they occur. In this way we can exploit the potential benefits of visual presentations in Web surveys, while avoiding the more obvious pitfalls.

3. Research Design and Source of Data

In order to explore some of these issues, we embedded three image experiments in a larger survey containing experiments on several other design issues. All experiments were fully crossed to avoid possible confounding or carryover effects. The instrument for the larger study contained items on a variety of topics culled from existing surveys. The instrument was programmed in SPSS's mrInterview by staff at MSInteractive. The sample was obtained from Survey Sampling, Inc. (SSI). SSI has two Web sample sources:

- SSI's Survey Spot sample is an opt-in Web panel of almost one million persons who have signed up on-line to receive survey invitations.
- SSI's eLite is an aggregated list of over seven million records from the opt-in lists of a wide variety of Web sites. It is not a panel in the usual sense, but a large aggregated list of persons who have given their permission to be contacted for surveys, promotional offers, advertising, and information of potential interest to users of the providing Web site.

We used sample from both sources. The standard SSI incentive (a sweepstakes drawing for cash prizes totalling $10,000) was used. Sample persons were invited to participate by email and nonresponders received one reminder. The URL directed sample persons to the MSInteractive website. The survey was conducted in March/April, 2003. The response rates for the two SSI samples are presented below:

Source	Number Contacted	Initial Responses	Completed Interviews	Response Rate
Survey Spot	7,838	1,555	1,379	17.6%
eLite	31,379	1,624	1,343	4.3%
Total	39,217	3,179	2,722	6.9%

We note that this is not a probability sample, but our goal is randomization (in an experimental sense), rather than representation. All of the analyses below are based on the completed cases only (n=2,722), but the results do not change if partials are included. We describe each of the three experiments and present the results in turn below.

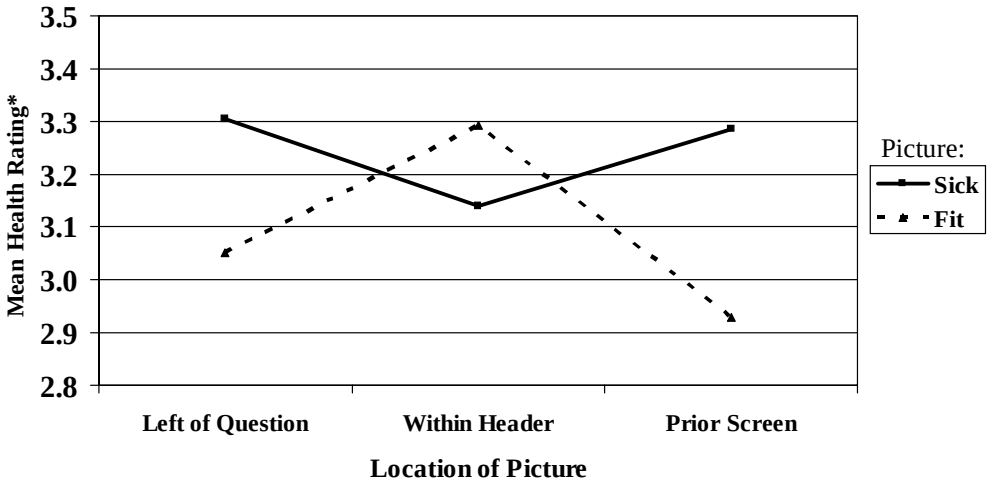
Experiment 1

The first experiment was designed to explore whether images could generate contrast effects, by suggesting an extreme standard for a judgement. Specifically, the target question was a standard self-

rated health item, “How would you rate your health?” The experimental manipulation consisted of a 2*3 design, with two different photographs (a woman jogging, and a woman lying in a hospital bed) crossed with three different locations (to the left of the question text, in the header, or on the previous screen). A seventh no-picture condition was added.

The pictures did indeed affect the self-ratings of health, lowering them on average for the respondents who got the picture of the healthy woman jogging (mean=3.36, s.e.=0.027) and raising them on average for those who got the picture of the sick woman in bed (mean=3.42, s.e.=0.027), with high numbers indicating better health. The overall effect of the picture was only marginally significant (*F*=3.08, d.f.=1, 2339, p<.08). To our surprise, there was also a significant interaction between the location of the picture and its content; that interaction is displayed in Figure 1. The overall model now reaches statistical significance (*F*=3.4, d.f.=5, 2332, p<.01) as does the interaction term. When the picture is in the header, assimilation (i.e., people’s ratings are more like that of the pictured woman) rather than contrast seems to be the result.

Figure 1: Health Ratings, by Picture and Location



* Scale range of 1-5; a high score indicates good health

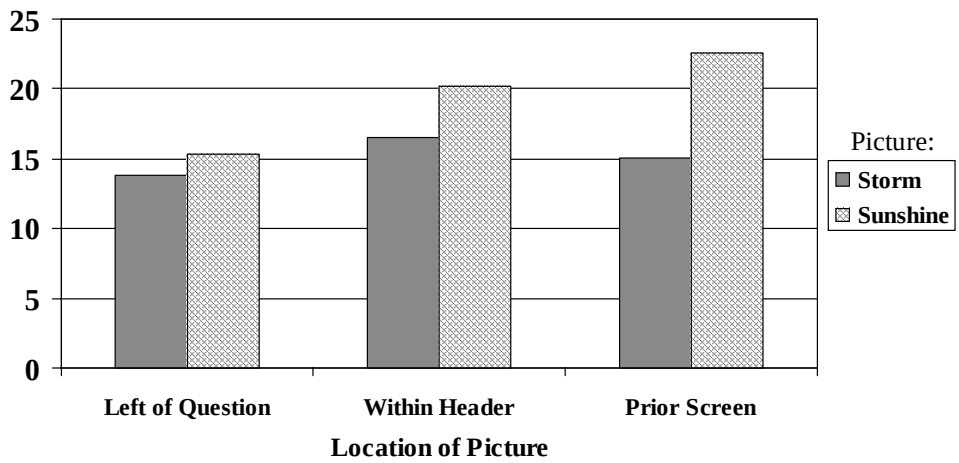
But perhaps more interesting, the picture experiment was crossed with another experiment in which both the spacing (visual position) and the labels (verbal position) of the response options were varied. While the spacing manipulation did not interact with the image manipulations, the wording of the response options did. Specifically, when the five response options ranged from “extremely good” to “good,” then “neutral,” “poor,” and “extremely poor,” the type of picture seen by the respondent appeared to have little effect on their answers. However, when the response options ranged from “excellent” to “very good, “good,” “fair,” and then “poor,” the effect of the picture is magnified. Specifically, under the “extremely good” to “extremely poor” condition, there is no significant difference in the mean health rating (*F*=0.34, d.f.=1, 1775, n.s.), but under the “excellent” to “poor” condition, the mean health rating for those seeing the picture of the fit woman was 3.09 compared to 3.24 for those exposed to the picture of the sick woman (F=7.61, d.f.=1, 1156, p<.01). Put in percentage terms, 33% of respondents report their health as excellent or very good when exposed to the picture of the fit woman, while 41% do so when exposed to the picture of the sick woman, a significant contrast effect. This leads us to posit that the greater the number of plausible responses available to the respondent—and the closer the points conceptually on the scale—the more susceptible the responses are to such contrast effects.

Experiment 2

Our second experiment focused on the priming of an affective state. Research by Schwarz and Clore (1983) found that people asked about their general happiness on a sunny day gave significantly more positive ratings than those asked on a rainy day. To what extent would this translate to pictures in a Web survey? To test this, the survey included a series of 9 items from two subscales (vitality and mental health) from the RAND 36-item health questionnaire (see Bell and Kahn, 1996). These items asked respondents how often in the past month they had felt "downhearted and blue," "full of pep," calm and peaceful," and so on. The picture manipulation consisted of a sunny sky or a stormy sky. This was again crossed with the same three location variations as Experiment 1. We expected the effect of the pictures to be largest in the condition where they were displayed on the prior screen, as to be effective, the prime must occur before the target question is displayed. The nine target items were displayed on two screens, with the order randomized across four sequences in a partial Latin square design.

The results reveal some intriguing patterns. First, the means of the two scales appear to be unaffected by the picture manipulation, either overall or in combination with the location variable. However, when looking at the individual items, we find that some appear to be more susceptible to the picture manipulation than others. Specifically, the item, "During the past 4 weeks, how often have you felt sad and blue" (arguably, an item than is most about mood), shows effects in the expected direction. Moreover, these effects are stronger when the item appears on the first of the two screens, and stronger still when the item is the first in the scale. The correlation between the type of picture and the response to this item increases from 0.012 when the item appears anywhere on the scale (n=2,362) to 0.032 when it appears on the first screen (n=1,769) and 0.056 when it appears as the first item (n=587). Furthermore, the effect of the image appears to increase with its location relative to the target question, with the difference between the two picture conditions smallest when the picture appears alongside the items, and largest when it appears on the prior screen. Consistent with the literature on stimulus onset asynchrony (SOA) and priming (e.g., Scharlau, 2002), this suggests that the effect of the image is optimal when it appears just before exposure to the target item, and when it is directly relevant to the target item (feeling sad and blue). This is illustrated in Figure 2 for the case where the item appears in the first position on the scale.

Figure 2: Percent Responding "None of the time" to the Item "During the past 4 weeks, how often have you felt sad and blue" by Picture and Location



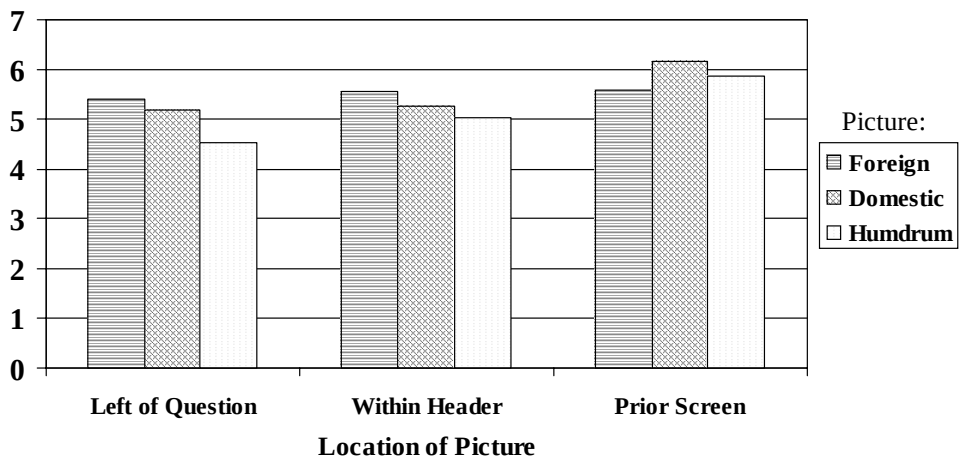
Again, while the overall effects may be relatively subtle, this exploratory research suggests that the presence of images can momentarily affect mood and consequently move responses in items susceptible to mood manipulations. The context in which such measures as depression scales are administered may well affect the outcomes.

Experiment 3

Our third experiment was a replication and extension of an earlier study (Couper, Tourangeau, and Kenyon, 2003). In that study, conducted using Knowledge's Network's WebTV-based panel, we found that showing respondents a picture of a high-frequency exemplar increased the reporting of related behavioural frequencies (shopping, travel, attending sporting events, etc.) over showing a lower-frequency exemplar of the same class of activity. Our current study focused particularly on travel, and sought to explore the effect of displaying a montage of several examples of the behaviour in question, i.e., overnight trips in the past year. Specifically, the montages contained either a set of foreign travel destinations (e.g., Eiffel Tower, Taj Mahal), domestic travel destinations likely to be reached by air (e.g., Statue of Liberty, Golden Gate Bridge), or a set of more mundane domestic images (e.g., highway signs, cornfield and barns) likely to evoke road trips. Our expectation was that foreign trips are less frequent than domestic trips by air, and both are less frequent than trips taken by car, and hence the montage of foreign destinations would yield the fewest trips reported and the mundane images the most. Again these were crossed with the three location conditions, for a 3x3 design, plus a tenth, no-picture, control condition.

In our initial pass at the analysis, we found no main effects, despite removing outliers and doing a log transformation of reported number of trips to account for the marked positive skew (as we did in the earlier study). The mean number of trips appear to be unaffected by the picture manipulation. However, the Poisson distribution is ideally suited for cases such as this, where the dependent variable is a count of events (number of trips), the variable is skewed and non-negative, and the variance increases as the mean increases. We thus ran a Poisson regression model, with picture content, location, and their interactions as predictors. The model is statistically significant ($G=92.119$, $d.f.=15$, $p<.001$), and both main effects and the interactions are significant. The estimated means from the Poisson regression are presented in Figure 3.

Figure 3: Poisson Regression Estimates of Mean Number of Trips By Picture and Location



This experiment yielded several surprises. First, the overall effects of the travel montages on the number of trips taken, and on follow-up questions on the duration, purpose and mode of transport of

the last trip, were modest—much less powerful than the earlier experiment using single exemplars. Second, the direction of the effect seems to be contrary to what we expected, with the humdrum picture yielding lower estimates of number of trips than the more exotic domestic locations and (in two of the three location conditions) the foreign destinations yielding the highest estimates. One *post hoc* explanation may be that the montages were simply too small overall, and thus contained too much detail, for respondents to sufficiently discriminate the content of the images, thereby reducing the differences among the conditions. Winkielman et al. (2003) introduce the notion of visual fluency, and suggest that the easier the image is to process, the stronger the effect of the image on the responses (see also Reber and Schwarz, 1999). The relatively small size of the images introduced in this experiment, along with the fact that each image contained four different scenes, may have reduced their effectiveness.

4. Discussion and Conclusions

We have briefly discussed several experiments on the use of images in Web surveys, and described preliminary results. Several broad observations can be made on the basis of these results. First, the inclusion of images in Web surveys can and does have an effect on the answers provided. Even images seemingly unrelated to the content, such as those displayed on a lead-in screen or appearing in the header area, can affect responses. Second, the effects are complex. As suggested at the outset, the effect of images on measurement error is a complex one, depending on many factors, related not only to the characteristics of the image (content, size, etc.) and the target question, but also to the juxtaposition of the two in the survey. In our experiments, we found some support for hypothesized effects, but also some unexpected results.

Clearly, there is much more work to be done to understand the role of images and other visual embellishments in shaping respondents' answers in Web surveys. But for now our message to practitioners must be to use images with caution. Over several experiments we have found no positive effects of images on motivating respondents to complete the survey, i.e., reducing abandonments. To be sure, there are many valuable applications of visual enhancements in Web survey measurement, but in the case where the images are not germane to the measurement, for now we believe the dangers of including images simply for the purpose of adding visual appeal may outweigh the benefits.

References

Bell D.S., and Kahn, C.E. (1996), "Assessing Health Status Via the World Wide Web." In J.C. Cimino (ed.), Proceedings of the AMIA Annual Fall Symposium. Philadelphia: Hanley and Belfus, pp. 338-342.

Couper, M.P., Tourangeau, R., and Kenyon, K (2003), "Picture This! An Analysis of Visual Effects in Web Surveys." Public Opinion Quarterly, forthcoming.

Duchastel, P.C. and Waller, R. (1979), "Pictorial Illustration in Instructional Texts." Educational Technology, 19 (11): 20-25.

Horn, R.E. (1998), Visual Language; Global Communication for the 21st Century. Bainbridge Island, WA: MacroVU, Inc.

Huddy, L., and Gunnthorsdottir, A.H. (2000), "The Persuasive Effects of Emotive Visual Imagery: Superficial Manipulation or the Product of Passionate Reason?" Political Psychology, 21 (4): 745-778.

Klauer, K.C. (1998), “Affective Priming.” In W. Stroebe amd M. Hewsatone (eds.), European Review of Social Psychology, Volume 8. New York: Wiley.

Lohr, L.L. (2000), “Designing the Instructional Interface.” Computers in Human Behavior, 16: 161-182.

Macrae, C.N., Bodenhausen, G.V., Milne, A.B., and Calvini, G. (1999), “Seeing More Than We Can Know: Visual Attention and Category Activation.” Journal of Experimental Social Psychology, 35: 590-602.

Prior, M. (2002), “More Than a Thousand Words? Visual Cues and Visual Knowledge.” Paper presented at the annual meeting of the American Political Science Association, Boston, August.

Reber, R., and Schwarz, N. (1999), “Effects of Perceptual Fluency on Judgments of Truth.” Consciousness and Cognition, 8: 338-342.

Scharlau, I. (2002), “Leading, but Not Trailing, Primes Influence Temporal Order Perception: Further Evidence for an Attentional Account of Perceptual Latency Priming.” Perception and Psychophysics, 64 (8): 1346-1360.

Schwarz, N. and Clore, G.L. (1983), “Mood, Misattribution, and Judgments of Well-Being: Informative and Direct Functions of Affective States.” Journal of Personality and Social Psychology, 45: 513-523.

Schwarz, N. (1994), “Judgment in a Social Context: Biases, Shortcomings, and the Logic of Conversation.” Advances in Experimental Social Psychology, 26: 123-162.

Schwarz, N. and Sudman, S. (1992), Context Effects in Social and Psychological Research. New York: Springer-Verlag.

Tourangeau, R, and Rasinski, K. (1988), “Cognitive processes underlying context effects in attitude measurement.” Psychological Bulletin, 103, 299-304.

Winkielman, P., Schwarz, N., Reber, R., and Fazendeiro, T. (2003), “Cognitive and affective consequences of visual fluency: When seeing is easy on the mind.” In L. Scott and R. Batra (Eds.), Persuasive Imagery: A Consumer Response Perspective. Mahwah, NJ: Erlbaum, pp. 75-89.

Zillman, D., Knobloch, S., and Yu, H. (2001), “Effects of Photographs on the Selective Reading of News Reports.” Media Psychology, 3: 301-324.

About the Authors

Couper, Conrad, and Tourangeau have appointments in the Institute for Social Research at the University of Michigan and in the Joint Program in Survey Methodology at the University of Maryland. This study is part of a program of research on Web survey design funded by the National Science Foundation (Grant SES-0106222). We thank Reg Baker, Andrey Peytchev, and Jenefer Willem for their contributions to the design, execution and analysis of this study. Address correspondence to: Mick P. Couper, Institute for Social Research, P.O. Box 1248, Ann Arbor, MI 48106, USA; e-mail: mcouper@umich.edu.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Salutations and Response Rates to Online Surveys

Jerold Pearson & Roger A. Levine

1. Background

As online surveys become more feasible for more populations, researchers are becoming more interested in understanding various issues related to response rates in the online environment. One issue of particular interest deals with the salutation used in invitations to online surveys. There has been much speculation about whether the salutation might affect response…and, if so, what kind of salutations work best for which populations. But there is little reliable data from which to make inferences.

An experiment was therefore undertaken to test four different salutations in a survey of alumni from Stanford University conducted on the World Wide Web in September 2002. The survey explored issues dealing with the university's logos, image, and branding; and, as such, was salient to the general alumni population, and not just to certain sub-groups.

2. Methodology

The population was defined as people who earned an undergraduate degree from Stanford from 1955 through 2002, who live in the continental USA, and for whom Stanford has an e-mail address. Alumni with only a graduate degree from Stanford were not included in the population, nor were Stanford employees, spouses of employees, or alumni flagged "Do Not E-mail."

Four random samples of 800 alumni were selected from the population, and each of the 3200 individuals was e-mailed a short invitation to participate in the survey. In addition to explaining the purpose of the study, its importance to the university, its salience to them as alumni, and its brevity, the invitations included a link to the survey (hosted on a Stanford web server) and a unique Respondent ID number. The invitations were exactly the same for all 3200 alumni – except for the ID number and the salutation:

- Sample 1 received a generic salutation ("Dear Stanford Alum").
- Sample 2 received a familiar personalized salutation ("Dear James").
- Sample 3 received a familiar personalized salutation without the "dear" ("James").
- Sample 4 received a formal personalized salutation ("Dear Mr. Bond").

The invitations were e-mailed on September 12, 2002, and two reminders were e-mailed to non-respondents before the survey was taken off the web on October 9, 2002.

Names, titles (ie. Mr, Dr, Ms, Professor, Reverend, Major General, etc), and demographic data (class year, gender, school, donor status, and Stanford Alumni Association membership) for each person in the four samples came from Stanford's alumni database.

"Donors" were defined as those alumni who had made a gift of any amount to any Stanford designation in Fiscal Year 2002 (September 1, 2001-August 31, 2002).

For the analysis of this data, undergraduate class year was used as a proxy for age:

- Class year 1994-2002 approximately correspond to alumni under 30.
- Class year 1975-1993 approximately correspond to alumni 30-49.
- Class year 1955-1974 approximately correspond to alumni 50 and older.

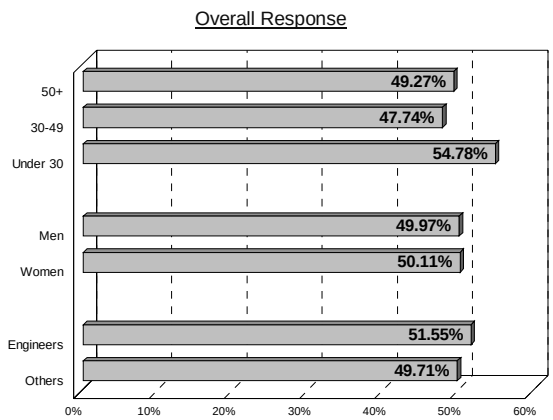
The random samples were proportionally representative of the entire alumni population (as defined above):

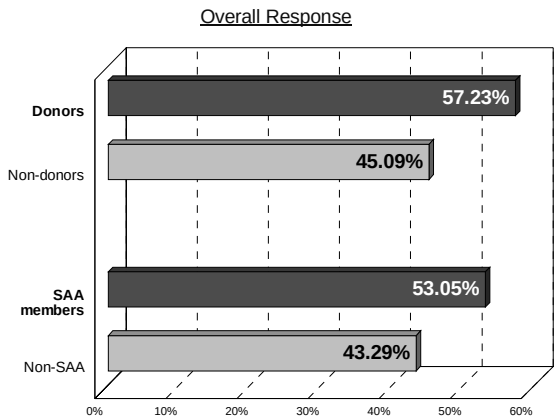
- 57% were men and 43% were women.
- 26% were under 30, 43% were 30-49, and 31% were 50 or older.
- 41% were donors and 59% were non-donors.
- 69% were Stanford Alumni Association members and 31% were non-members.
- 79% had undergraduate degrees from the School of Humanities and Sciences, 18% from the School of Engineering (which includes computer science and other computer-related departments), and 3% from the School of Earth Sciences.

3. Results

Overall Response

Only 92 of the 3200 e-mail addresses (2.87%) were invalid. Of the 3108 valid addresses, 1555 completed the survey (50.03%).





Current research among Stanford alumni indicates that women, non-engineering alumni, and older alumni are as likely as men, engineers, and younger alumni to have access to the web – but they *use* the web less frequently. So there is reason to be concerned that response rates to online surveys might be lower among these groups. Nevertheless, overall response (to all four samples combined) did not differ by gender or school (engineers vs non-engineers). And, while the response rate was a few percentage points higher among younger alumni, the difference is not quite statistically significant.

Not surprisingly, however, overall response was significantly greater among alumni with whom the university has the strongest relationship. Response was greater among:

- Donors (57.23%) than non-donors (45.09%). [$p < .0001$]
- Stanford Alumni Association members (53.05%) than non-members (43.29%). [$p < .0001$]

Response to the four salutations did not significantly differ overall; no salutation drew a significantly greater response than the others did:

- 47.87% for "Dear Stanford Alum."
- 50.26% for "Dear James."
- 50.25% for "James."
- 51.77% for "Dear Mr. Bond."

Generic vs Personalized Salutations

Overall response to the three personalized salutations taken as a whole (50.75%) was about 3 percentage points higher than response to the generic salutation (47.87%), but the difference is not statistically significant. Similarly, within most of the demographic groups, response was a few percentage points higher for the personalized salutations – but the only statistically significant differences are among:

- Alumni 50 and older – 51.19% of whom responded to the personalized salutations, compared to only 43.46% who responded to the generic salutation [$p < .05$].
- Alumni who are *not* members of the Alumni Association – 45.26% of whom responded to the personalized salutations, compared to only 37.45% who responded to the generic salutation [$p < .05$].

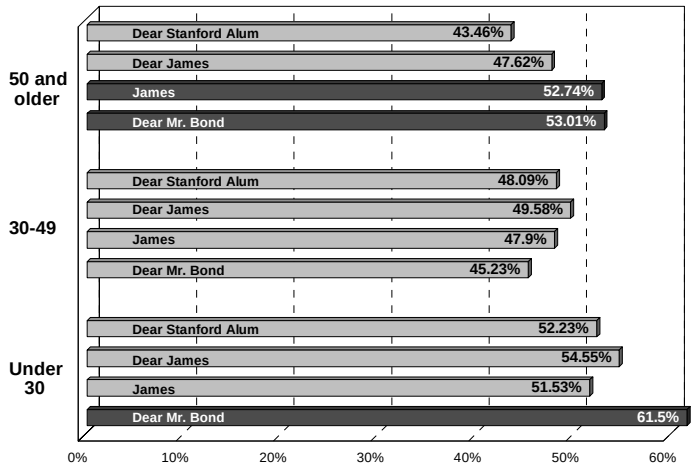
It may seem surprising that the personalized salutations made more of a difference with non-members (with whom Stanford has a *less* personal relationship) than with members of the Alumni Association. Upon reflection, however, a plausible explanation comes to mind. As noted, Alumni Association members were more likely to respond overall; perhaps their closer ties to the university help elicit a similar response rate regardless of the salutation. Non-members, on the other hand, may feel that a personalized salutation suggests that the university still recognizes them and still considers them part of the community, and they might respond favorably to that.

The Full Monty: All Four Salutations

While there were not many significant differences in response to the generic vs the personalized salutations (as a whole), more differences are revealed when we take a more granular look at response to each of the four salutations. In other words, among certain demographic groups, some personalized salutations elicited a greater response than others did.

Age: Among alumni 50 and older, the formal salutation (Dear Mr. Bond) and the familiar salutation without the "dear" (James) drew greater responses than the generic salutation did. Perhaps "older" alumni appreciate the respect that the formal salutation may imply. If so, their similar response to the familiar salutation is baffling.

Response Rates to the 4 Salutations by Age



- Dear Stanford Alum: 43.46%
- Dear James: 47.62%
- James: 52.74% [p < .05]
- Dear Mr. Bond: 53.01% [p < .05]

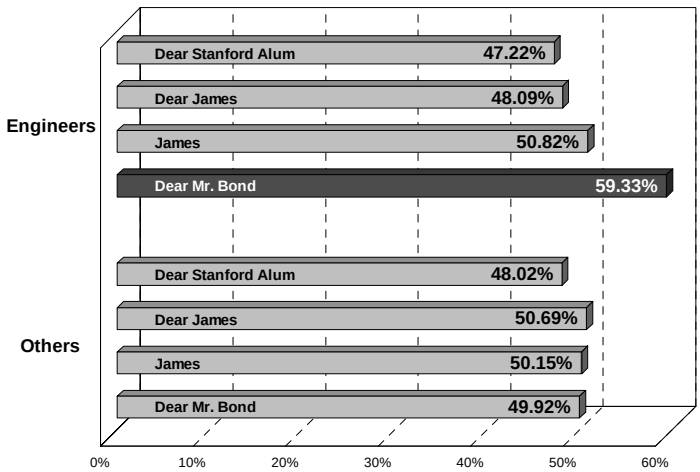
No salutation drew a significantly greater response than the others did among alumni age 30-49.

Alumni under 30 also appear to have responded better to one of the personalized salutations than they did to the generic salutation. But, perhaps surprisingly, it is the *formal* – not one of the familiar – salutations that seems to have elicited the best response from them (although, because the samples are somewhat smaller, the difference is just a hair shy of statistical significance). Again, younger alumni may appreciate the respect implied in the formal salutation.

- Dear Stanford Alum: 52.23%
- Dear James: 54.55%
- James: 51.53%
- Dear Mr. Bond: 61.50%

School: The formal salutation (Dear Mr. Bond) elicited a greater response than the other salutations did among alumni from the School of Engineering.

Response Rates to the 4 Salutations by School



- Dear Stanford Alum: 47.22%

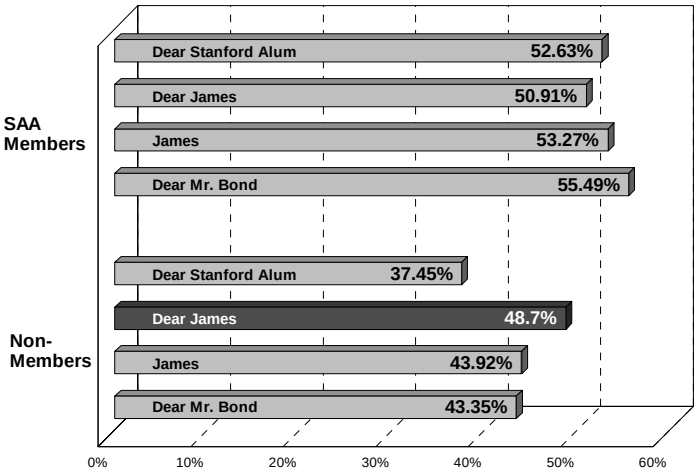
- Dear James: 48.09%
- James: 50.82%
- Dear Mr. Bond: 59.33% [p < .05]

No salutation drew a significantly greater response than the others did among non-engineers (alumni from the School of Humanities and Sciences and the School of Earth Sciences).

Stanford Alumni Association Membership: No salutation drew a significantly greater response than the others did among Alumni Association members.

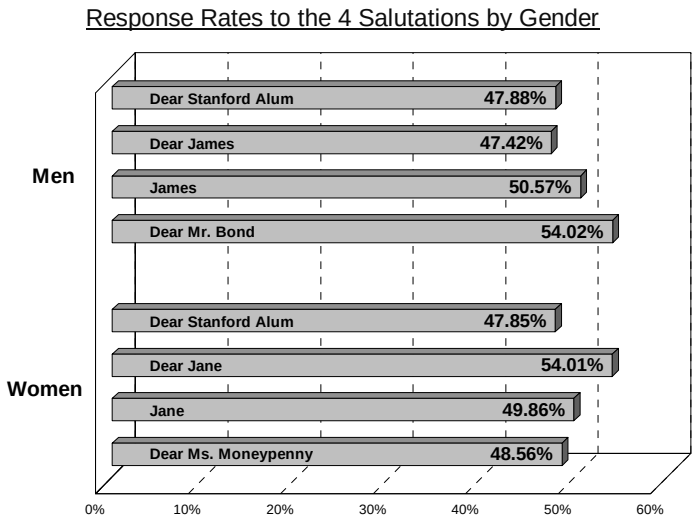
But among non-members, the familiar salutation (Dear James) drew a significantly greater response than the generic salutation did, and a few percentage points more than the other personalized salutations did.

Response Rates to the 4 Salutations by SAA Membership

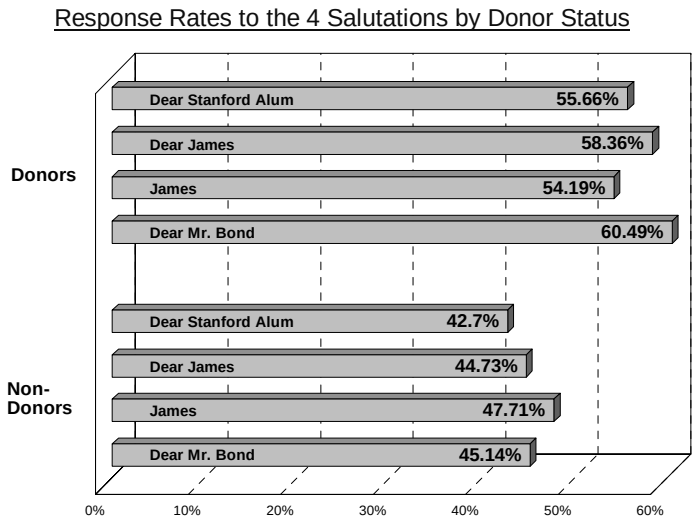


- Dear Stanford Alum: 37.45%
- Dear James: 48.70% [p < .05]
- James: 43.92%
- Dear Mr. Bond: 43.35%

Gender: Men responded a few percentage points better to the formal salutation (Dear Mr. Bond) than to any other…while women responded a few percentage points better to the familiar salutation (Dear Jane). Nevertheless, no salutation drew a *significantly* greater response than the others did from either men or women.



Donors and Non-donors: No salutation drew a significantly greater response than the others did from either donors or non-donors – although, among donors, the formal salutation (Dear Mr. Bond) outdrew the others by a couple of percentage points.



4. Summary and Implications

As expected, if the sponsor of a survey is identified upfront to respondents (as was the case with this Stanford survey), response will be higher among people with whom the sponsor has the closest relationship (in this case, donors and Alumni Association members). This will hold true regardless of the salutation used.

Although women and non-engineers may use the web less than men and engineers do, this survey indicates that they are just as likely to respond to a web survey (assuming the survey is of equal salience to them). However, response to a web survey may be a few percentage points higher among "younger" than "middle-aged" and "older" people, although the difference may not be statistically significant.

Response was a few percentage points higher to the personalized salutations than to the generic salutation – both overall and within most demographic groups. So if the added time, effort, and

expense are not too onerous, it may be worth personalizing salutations when possible. There certainly appears to be no downside to personalizing (in terms of response rates). On the other hand, most of these differences were not statistically significant, so if personalizing is too difficult or not possible, overall response may not suffer much with a generic salutation.

Nevertheless, within certain demographic groups, some personalized salutations worked better than others, and response rates *were* significantly higher – suggesting that surveys conducted among specific homogeneous populations may benefit from personalization:

- Surveys conducted among older and (more surprisingly) younger people may net the highest response if they use a formal personalized salutation (Dear Mr. Bond).
- Surveys conducted among engineers (and perhaps by extrapolation) technically oriented people may also net the highest response if they use a formal personalized salutation (Dear Mr. Bond).
- Perhaps counter-intuitively, surveys conducted among people with whom the sponsor has a weaker relationship (ie. non-members) may net the highest response if they use a *familiar* personalized salutation (Dear James).
- Again counter-intuitively, the salutation did not seem to matter to Alumni Association members, so surveys conducted among a sponsor's closest constituents or customers may have less to gain from personalized salutations.
- Although the differences were not quite significant, it's possible that surveys conducted with men may net the highest response if they use a formal personalized salutation (Dear Mr. Bond)…while surveys conducted with women may net the highest response if they use a *familiar* personalized salutation (Dear Jane).

Of course, surveys conducted with diverse or heterogeneous populations may also benefit from personalization if the demographic characteristics of each person in the sample can be identified and the salutation is varied accordingly. Using multiple personalized salutations, however, may entail even more time, effort, and expense.

It is worth repeating that this survey was conducted exclusively with degree holders from Stanford University, who are more highly educated than many survey populations are. Since some response effects are less pronounced among highly educated respondents, the results of this experiment may differ with the population.

Because this work was exploratory in nature, a large number of subgroup comparisons (about 72) were performed. Under these circumstances, one might expect about three or four Type I errors (spuriously significant results at $p<0.05$), so it is possible that these significant results are due to chance. Therefore, it is strongly recommended that this experiment be replicated – especially if it can be replicated with different populations to see if the results vary by education or other characteristics.

About the Authors

Jerold Pearson, Director of Market Research, Stanford Alumni Association, Stanford University

Roger A. Levine, PhD, Managing Research Scientist, Director, Cognitive Survey Laboratory, American Institutes for Research

Appendix A: Response Tables

Total Sample

	Valid Addresses	Completes	Response %
All	3108	1555	50.03%
Men	1781	890	49.97%
Women	1327	665	50.11%
Under 30 (1994-2002)	805	441	54.78%
30-49 (1975-1993)	1349	644	47.74%
50+ (1955-1974)	954	470	49.27%
Donors	1265	724	57.23%
Non donors	1843	831	45.09%
Engineers	547	282	51.55%
Others	2561	1273	49.71%
SAA members	2147	1139	53.05%
Non members	961	416	43.29%

Dear Stanford Alum (generic salutation)

	Valid Addresses	Completes	Response %
All	775	371	47.87%
Men	449	215	47.88%
Women	326	156	47.85%
Under 30 (1994-2002)	224	117	52.23%
30-49 (1975-1993)	314	151	48.09%
50+ (1955-1974)	237	103	43.46%
Donors	309	172	55.66%
Non donors	466	199	42.70%
Engineers	144	68	47.22%
Others	631	303	48.02%
SAA members	532	280	52.63%
Non members	243	91	37.45%

Dear James

	Valid Addresses	Completes	Response %
All	782	393	50.26%
Men	445	211	47.42%
Women	337	182	54.01%
Under 30 (1994-2002)	198	108	54.55%
30-49 (1975-1993)	353	175	49.58%
50+ (1955-1974)	231	110	47.62%
Donors	317	185	58.36%
Non donors	465	208	44.73%
Engineers	131	63	48.09%
Others	651	330	50.69%
SAA members	552	281	50.91%
Non members	230	112	48.70%

James

	Valid Addresses	Completes	Response %
All	790	397	50.25%
Men	439	222	50.57%
Women	351	175	49.86%
Under 30 (1994-2002)	196	101	51.53%
30-49 (1975-1993)	357	171	47.90%
50+ (1955-1974)	237	125	52.74%
Donors	310	168	54.19%
Non donors	480	229	47.71%
Engineers	122	62	50.82%
Others	668	335	50.15%
SAA members	535	285	53.27%
Non members	255	112	43.92%

Dear Mr. Bond

	Valid Addresses	Completes	Response %
All	761	394	51.77%
Men	448	242	54.02%
Women	313	152	48.56%
Under 30 (1994-2002)	187	115	61.50%
30-49 (1975-1993)	325	147	45.23%
50+ (1955-1974)	249	132	53.01%
Donors	329	199	60.49%
Non donors	432	195	45.14%
Engineers	150	89	59.33%
Others	611	305	49.92%
SAA members	528	293	55.49%
Non members	233	101	43.35%

All Personalized Salutations

	Valid Addresses	Completes	Response %
All	2333	1184	50.75%
Men	1332	675	50.68%
Women	1001	509	50.85%
Under 30 (1994-2002)	581	324	55.77%
30-49 (1975-1993)	1035	493	47.63%
50+ (1955-1974)	717	367	51.19%
Donors	956	552	57.74%
Non donors	1377	632	45.90%
Engineers	403	214	53.10%
Others	1930	970	50.26%
SAA members	1615	859	53.19%
Non members	718	325	45.26%

Appendix B: Invitation and Reminders

Initial Invitation - September 12, 2002

From: Jerold Pearson (jpearson@stanford.edu)
Subject: Alumni Opinions Wanted

[salutation],

Stanford is currently taking a fresh look at some of the many logos it uses: How well do they reflect the image and qualities of the university, which ones have meaning to the Stanford community, and to what extent does the multiplicity of logos help or hinder effective communication?

To help us answer those questions and be responsive to alumni feelings, we are conducting a very brief – and, I think, fun – survey. You are included in the sample of alumni that has been randomly selected to participate in the survey. IT WILL ONLY TAKE ABOUT 3 MINUTES, and your thoughts and opinions will be instrumental to the review now underway.

In order to ensure the validity of the data, the survey is being conducted only with those of you who were randomly selected, so our efforts to listen to and serve alumni in this matter depend on your cooperation. In other words, I strongly encourage you to take the survey, which I think you will find enjoyable. All responses will be kept anonymous and strictly confidential.

Please click here [URL] to connect to the survey

and then enter this Respondent ID number: xxxx

which will verify that you are in the random sample. Thanks for your cooperation.

Jerold Pearson, '75
Director of Market Research
Stanford University
650-723-9186

1st Reminder - September 23, 2002

From: Jerold Pearson (jpearson@stanford.edu)
Subject: Your Thoughts About Stanford Logos

[salutation],

Just a quick reminder to please participate in the survey about Stanford logos. IT WILL ONLY TAKE ABOUT 3 MINUTES, and it's a fun survey. Alumni often tell us we should ask for their opinions...so we're asking! And your opinions on this issue really will matter; indeed, they will be instrumental to the review now underway.

How well do Stanford's logos reflect the image and qualities of the university, which ones have meaning to the Stanford community, and to what extent does the multiplicity of logos help or hinder effective communication?

You are included in the sample of alumni that has been randomly selected to participate in the survey, so our efforts to listen to and serve alumni in this

matter depend on your cooperation. All responses will be kept anonymous and strictly confidential.

To connect to the survey, please click here [URL]

and then enter this Respondent ID number: xxxx

which will verify that you are in the random sample. Thanks for your cooperation.

Jerold Pearson, '75
Director of Market Research
Stanford University
650-723-9186

2nd Reminder - October 2, 2002

From: Jerold Pearson (jpearson@stanford.edu)
Subject: Alumni Feedback: Last Call

[salutation],

In case you've been away or unable to participate until now, here's a final gentle reminder to please take the survey about Stanford logos. We really do want alumni guidance on this matter, and I guarantee that your opinions will factor into decisions that are made. IT WILL ONLY TAKE ABOUT 3 MINUTES.

In return for your time, I will send you the Executive Summary of my report after I've analyzed the data.

To connect to the survey, please click here [URL]

and then enter this Respondent ID number: xxxx

which will verify that you are in the random sample. Thanks for your cooperation.

Jerold Pearson, '75
Director of Market Research
Stanford University
650-723-9186

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Northern Ireland in 3D. Data, Dissemination and Discovery

Paula Devine & Katrina Lloyd

Abstract

ARK, the Northern Ireland Social and Political Archive, is a web resource based jointly at Queen's University Belfast and the University of Ulster. Its central aim is to exploit the web to make social and political material on Northern Ireland available to the widest possible audience. As such, its audience consists of academics, government policy makers, students, school children, the media, local voluntary groups and the general public. This necessitates information being available in a variety of formats, ranging from web pages through interactive searchable databanks and downloadable datasets at one end of a continuum to traditional printed papers at the other.

The ARK site (www.ark.ac.uk) provides a unified, searchable interface that allows access to a wide spectrum of material on the social and political life of Northern Ireland. Behind this interface is a range of resources which provide access to different types of both quantitative and qualitative information:

- Tables of results from academic social surveys.
- Survey data to download for secondary analysis.
- Bibliographies.
- Summaries of social policy research.
- Text of book chapters and articles.
- Visual information such as photographs of places and people.
- Chronologies.
- Election results.

Therefore, data, dissemination and discovery are key foci of ARK's work, in that

- Data can be disseminated in different ways, such as in tables or text.
- Data can be discovered by searching the entire archive, or by focusing on an individual component.

This paper will describe the experience of the ARK team in recent years, as well as describing its plans and ongoing programmes of work. The technical, financial, quality and contextual issues and problems encountered by developing and maintaining such an archive will be highlighted, as well as the need for such an archive to exist.

Keywords

Dissemination; Northern Ireland; Internet.

1. Introduction

ARK, the Northern Ireland Social and Political Archive, is a web resource based jointly at Queen's University Belfast and the University of Ulster. Its central aim is to exploit the web to make social and political material on Northern Ireland available to the widest possible audience. The underlying assumption of ARK is an archive that provides a bridge between the world of academia and the needs of ordinary people for information about their society. As such, its audience consists of academics, government policy makers, students, school children, the media, local voluntary groups and the general public. This necessitates information being available in a variety of formats, ranging from web pages through interactive searchable databanks and downloadable datasets at one end of a continuum to traditional printed papers at the other.

The ARK site (www.ark.ac.uk) provides a unified, searchable interface that allows access to a wide spectrum of material on the social and political life of Northern Ireland. Behind this interface is a range of resources which provide access to different types of both quantitative and qualitative information. However, we will begin by providing the history of ARK.

2. History

The origins of ARK are in the late 1990s, within two independent and successful web sites:

- CAIN.
- Northern Ireland Life and Times Survey.

CAIN

CAIN is the Conflict Archive on the INternet, which began in 1996, and is based at the University of Ulster. This is an encyclopaedic site providing information and source material on the Northern Ireland conflict and politics from 1968 to the present. More recently, CAIN has begun to provide information on general society in Northern Ireland. CAIN is a hugely successful site, and so far the site has attracted more than 13.5 million ‘hits’. Two thirds of its users come from outside Northern Ireland. The site has won many awards for example, BBC site of the week, and Encyclopaedia Britannica’s Internet Guide Site of the Day.

CAIN is a huge site, containing a wide variety of data, such as:

- Facts and figures.
- A bibliography of the conflict.
- Various chronologies within CAIN include a list of events before and during the Peace Process.
- Information about key events in Northern Ireland’s history, such as Bloody Sunday, the Anglo Irish Agreement, and the Peace process.
- A section on key issues including law and order, education and employment. For example, the section on Children and Young People in Northern Ireland examines their involvement in the Troubles and their roles as victims, perpetrators and peacemakers. The section includes relevant articles, statistics, and photographs.
- Visual material consisting of maps and murals, as well as photographs of places and people.
- Databases, such as deaths due to the Troubles; TV and films based on Northern Ireland.
- Abstracts on organisations, from the Ancient Order of the Hibernians to the Orange Order.
- Election results.
- Biographies of people prominent during the Troubles.

Northern Ireland Life and Times Survey

The Northern Ireland Life and Times (NILT) Survey was launched in the autumn of 1998. Its mission is to monitor the attitudes and behaviour of people in Northern Ireland to provide a time-series and a public record of how attitudes and behaviour develop on a wide range of social policy issues.

Life and Times is a direct descendent of the Northern Ireland Social Attitudes Survey (NISA) which ran from 1989 to 1996. NISA was a sister survey to the British Social Attitudes Survey (BSA), and, by running the same modules as BSA, it provided a time-series of social attitudes allowing comparisons with Britain. Against the background of the new political arrangements in Northern Ireland it was agreed that the new Northern Ireland Life and Times Survey would be better served by cutting its links with its British counterpart. NILT now carries on the tradition of a time series of attitudes but has shifted the focus away from comparisons with Britain. It is largely Northern Ireland focused, it is social policy focused, and it is designed to be used by the wider public in Northern Ireland. Nonetheless, every year includes a substantial component which either continues an old NISA time series, or replicates a BSA module. The Life and Times Survey is part of the International Social Survey Programme, which provides the scope for wider cross-national analysis.

Life and Times is run annually with a modular format. The range of modules included in the survey varies by year, although all the modules designed to be repeated in years to come.

However, two of these modules – Community Relations and Political Attitudes – are included every year. In fact, some community relations questions have been carried forward from the Northern Ireland Social Attitudes Survey, and so we are able to track attitude change since 1989. In a rapidly changing social and political environment such as Northern Ireland, it is essential to have a mechanism for charting shifts in attitudes and values to important topics such as national identity, integrated education and cross-community contact.

Approximately 1800 respondents aged 18 years or over are interviewed each year, primarily using CAPI. Respondents are also asked to complete a self-completion questionnaire on paper.

The survey aims to provide:

- A local resource for use by the general public.
- A data source for a more theoretical academic debate.

Thus, one important aspect of Life and Times is that it is a free and easily accessible resource. Integral to this strategy is our web site, which provides easy access to:

- All the questionnaires.
- Technical reports, providing information on the background to the survey, survey and sampling design, response rates, sampling errors and confidence intervals, and a comparison of demographic details with other surveys for example, the Continuous Household Survey and the Census.
- Publications, such as Research Updates on specific topics. So far more than 20 of these have been produced. These are available on the web site in PDF format, but we also produce them in paper format. We also continue the series of edited volumes ‘Social Attitudes in Northern Ireland’.
- Information on the helpline, which users can contact by phone or email if they require specific help or information.
- One key part of the web site is the production of tables of results for every question, broken down by age, sex and religion. The results for each year's survey are available on our web site in the following June, which is a quick turnaround for an academic survey.

- The raw data files in SPSS portable format, which can then be imported into any version of SPSS on any platform, whether it is PC, Macintosh or mainframe based. Thus, users can undertake secondary analysis themselves.

In the first year of the survey, the tables of results were created manually - all 500 of them - and took two people approximately six weeks to complete. However, as the survey progressed, it was recognised that this was not a feasible solution due to the amount of time involved, the potential for error, and, not least, the sheer tedium of the task. In order to automate the process, a specialised programme was written. This programme incorporates SPSS data, HTML templates and a manually created database of variable names and associated question text to create the tables. Once the database of variable names and text is prepared, it takes approximately six minutes to produce what used to take six weeks. We have used this system since 1999, and it has proved to be an invaluable tool.

This programme is still under development. Future plans include allowing users to perform tabulations in real time and plans to integrate it with a graphics package to allow production of graphs and other visual aids.

From 1998 to 2000, the Young Life and Times Survey recorded the views of all young people aged 12 to 17 years living in the same household as adult respondents. This survey included a subsection of the adult questions, as well as a module of particular importance to this age group, for example, Rights of the Child and Education.

The NILT site is accessed regularly by a large variety of users, from secondary school classes through journalists to academics and civil servants. We hope to develop the site to meet the needs of individual user groups. For example, we have just completed a pilot project with teachers attending a Masters course in the Graduate School of Education in Queen's. We created an online survey based on a subsection of the Young Life and Times Survey, which teachers' classes completed. We then compared these results with results from Young Life and Times respondents. The differences and similarities between results made for an interesting discussion, and we will be undertaking workshops with the participating classes next term.

3. Integration

In 2000, the team was successful in obtaining funding for a third project: the Online Research Bank (ORB). This provides a searchable online bibliography of social policy based research that has been undertaken in Northern Ireland since 1990. Where possible, a short summary of the research is written, or obtained from the authors. A short background to the research, the methodology and the key findings are presented. The bibliography now contains more than 1000 references, with summaries, or links to online summaries for more than half of these entries. Even in a small place like Northern Ireland, much research is undertaken, but not made publicly available. Thus, ORB facilitates public access to the world of academic, voluntary sector and government social policy research and offers summaries as a means of disseminating information to those who do not have the resources to access the original documents. ORB is gradually establishing a reputation as an important resource for community organisations and the public sector.

The development of ORB meant that three related resources were being managed by a dispersed group of researchers. The next logical step was to merge these resources into a social and political archive. In mid 2000, funding was obtained for a year to run a pilot.

At the end of the pilot, a successful evaluation was undertaken which showed in-depth knowledge, interest and use of the individual ARK components across a wide spectrum of users. Despite this, the overarching concept of ARK was not especially widely known. This reflected our own views on progress to date. Although the three resources were all under the umbrella of ARK, functionally the only links between them was a common server for NILT and ORB, as well as hyperlinks between the sites.

Since 2001, developments within ARK have taken place on two levels. Firstly, new resources have been added. Secondly, we have been focusing on building ARK as a well-known, unified, searchable resource.

4. New resources

Survey Analysis Unit

The Survey Analysis Unit was created in recognition that many potential users of social survey data need serious technical support. So far, the Survey Analysis Unit has attracted a wide variety of clients, including the Equality Commission for Northern Ireland, the Rural Development Council, the Department of Health, Social Services and Public Safety as well as the two universities in Northern Ireland. It is encouraging that we have attracted these contracts without ever advertising our existence, thus proving that there is a need for analysis support.

Surveys Online (SOL)

SOL aims to improve take-up of the many large-scale social surveys already carried out in Northern Ireland by making them publicly available on-line. SOL follows the successful Life and Times strategy of providing tabulations and cross-tabulations of the responses to the questions in each survey. These are produced using the same specialised programme written to create Life and Times tables of results. One component of SOL is a time series of attitudes to Community Relations issues from 1989 to the present

Northern Ireland Household Panel Survey (NIHPS) Analysis Unit

A recent development within ARK is a unit dedicated to facilitating user access to the new NIHPS - an extension of the long-running British Household Panel Survey.

Northern Ireland Elections web site

Northern Ireland Elections is an essential international resource for politicians, researchers, the media, students and anyone interested in Northern Irish elections and their outcomes. It provides information and results on all parliamentary, Assembly, local and European elections since 1973.

Young Life and Times (YLT)

From 1998 to 2000, a Young Life and Times Survey ran alongside the Life and Times Survey. As previously mentioned, all young people aged between 12 and 17 in the same household as an adult respondent were interviewed. This resulted in approximately 500 respondents. As with the adult survey, tables of results, as well as the raw data, were available from the web site. However, after 2000, a review of the need, content and methodology of Young Life and Times was carried out. This was partly due to low response rates.

The review concluded that there was unanimous support for a social attitudes survey of young people that is independent from the adult survey. Such support was particularly enthusiastic from those in the voluntary sector who deal with social issues and policy affecting young people. However, the 12 to 17 years age group was viewed as being inappropriate. The relevancy of topics and the appropriate question wording for the different ages was seen as being too wide to allow for a single questionnaire. We have now obtained funding for two years of YLT, focusing only on young people aged 16 years. The content will also be specific to this survey, and not dependent on the adult Life and Times Survey. In this pilot year, we are experimenting with a mixed mode approach. Respondents will be able to choose to complete the survey on the phone, on paper or via the web. Results will be made available in late autumn 2003 using a variety of methods, including the web site, the public launch of two Research Updates and a continuation of the project using qualitative research methods.

5. ARK as a unified resource

As well as developing new resources, ARK has developed functional integration between them. A new 'front end' has been created, which allows searching across all resources, using a defined set of keywords, for example, conflict, equality issues, transport and community relations.

Results are then displayed showing a brief description of the material and the URL, as well as the type of information available. The type of information provided is split into six categories:

1.Written Materials

2. Research summaries

3. Audio visual

4. Election results

5. Survey results

6. Facts and Figures

6. Technical issues

The first step in creating a unified resource was to load the sites onto one server. ARK is a project split across the various campuses of the University of Ulster and Queen's University. The separate components of ARK were also spread across different servers within the two universities. Having secure funding for and purchased a new server, the component parts of ARK where transferred.

The next step was to integrate the component parts and provide a mechanism that searched all the parts seamlessly in one click but did not produce reams of superfluous results often produced by the most popular search engines. It was important to produce focused results that pointed to key pages across the whole site.

With this in mind we created an ARK database. This contains information on subject, URLs, descriptions and information types for the key pages within each site. It was not necessary to do this for ORB as it was already based on a similarly structured database and we could tie the two together using the subject field. An entry form was produced to populate the database and a search program was then implemented that allowed the user to search on a list of predetermined subject types across the two databases. For example, if we search on housing, the results from Life and Times show just the index page for the 1998 Housing module, as well as a relevant Research Update. It does not produce a list of the 50 pages containing tables of results within the module.

7. Dissemination

Dissemination from ARK is multi faceted. Obviously, being predominantly web based, we register with search engines and resource gateways, such as the Social Science Information Gateway (SOSIG: http://www.sosig.ac.uk). However, more traditional methods are utilised. We run an annual series of seminars which has been well attended. These are advertised on the web site, on posters and by email.

Publications, such as short Research Updates and Occasional Papers are available on the web site in PDF format, as well as in paper form. These are distributed to libraries, organisations and individuals on our mailing list. We also produce a book on Social Attitudes in Northern Ireland. Despite all our best efforts to provide a wealth of information on the web, some people still prefer to use books.

We also run workshops on particular topics. For example we worked with the Rural Development Council to run a one day workshop on commuting and transport, focusing on data issues. A series of workshops on the use of the Northern Ireland Household Panel Survey will begin this autumn.

8. Quality

As with any project or resource, our reputation is only as good as the data we disseminate. Within the Life and Times Survey we operate the usual good practice quality controls, such as data validation, and confidentiality checks. The fieldwork is undertaken by an outside agency, whose field force operate to the Interviewer Quality Control Scheme (IQCS). Ten per cent of interviews are back checked by fieldwork managers, and this involves telephone back checking as well as personal visits to respondents' homes. The fieldwork agency is also registered under the Data Protection Act.

Within Surveys Online, only methodologically robust academic surveys are included. For both of these resources, our software will only produce tables of results if the cell size is above a certain minimum level.

All entries on ORB are checked and verified for accuracy before being loaded onto the database.

9. Problems encountered by developing and maintaining such an archive

One of the main problems encountered was that ARK originally began with separate web sites. These were on different servers, designed and maintained by different people. Bringing these together has been a logistical nightmare that we are still working on. Obviously, creating the searchable database was very time consuming. Redesigning the individual components so that they have a comparable 'look' may never be possible due to time constraints of staff who are working hard just to maintain the information within the sites.

Another important issue is that of the URL. When Life and Times was on a Queen's University server, it obviously had a Queen's web domain. CAIN was located on a University of Ulster server, and so had their web domain. However, it was very important for the ARK team to have web addresses that did not give allegiance to just one particular university. We applied for, and fortunately were successful in obtaining a separate URL which is independent of the two universities: www.ark.ac.uk

Although Northern Ireland is a small place, and thus a limited amount of research is carried out here, there are still problems in finding out what has been undertaken in the past. In fact, that is one of the reasons why ARK was created in the first place.

Getting access to such information is also sometimes difficult, although the internet has made it easier to find out about and access some types of information.

10. Need for such an archive to exist

The main reason for creating ARK – to provide access to social and political information – is as pertinent now as it was five years ago, and is likely to remain that way. There is a need to provide access to information, and also it is beneficial to provide that information in one umbrella site. The new searching facility within ARK, meaning that users can search across all constituent components provides a much-needed link between all the embedded sources of information. We also have evidence that data providers appreciate the benefits of such an archive. Instead of having to publicise ARK and search hard and wide for information, we now find that we are beginning to be offered material. This can take the form of researchers sending us information on their work for inclusion in ORB or for inclusion in Surveys Online. It also takes the form of researchers offering us their complete project, or in one case, a complete site, for inclusion for ARK. After checking the relevance and quality of these resources, we were happy to include them in ARK.

11. Conclusion

In conclusion, data, dissemination and discovery are key foci of ARK's work:

- Data can be disseminated in different ways, such as in tables or text.
- Data can be discovered by searching the entire archive, or by focusing on an individual component.

In this way, we can view Northern Ireland in 3D!

About the Authors

Paula Devine and Katrina Lloyd are both Research Directors of ARK and are based in the Institute of Governance, Public Policy and Social Research at Queen's University Belfast. ARK is the Northern Ireland Social and Political Archive, which makes material on the social and political life of Northern Ireland available to the widest possible audience (www.ark.ac.uk). Paula Devine can be contacted at p.devine@qub.ac.uk, and Katrina Lloyd's email address is k.Lloyd@qub.ac.uk

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Successfully Delivering Technology for Online Research

Leonard Bayer & Peter Milla

Abstract

With the Internet revolution approaching its adolescent stage, the research industry in the United States expects growth in Internet research to be in excess of 40% for the next several years. With Europe quickly coming to Internet penetration levels achieved in the US, it is expected that forward thinking research industry agencies in the UK and Europe will be the aggressively moving their clients' work to the Internet.

With the decision to move work to the Internet, research agencies will have the need to provide or procure the technology required for online research. This paper will address the various technology areas that are a necessary part of a successful solution, including:

- Sampling and Panel Management
- Load Balancing, Authentication, Survey Re-entry
- Scalability and Reliability
- Security
- Survey Functionality
- Deployment and Re-deployment
- Management/Quota Management
- Aggregation, Data Weighting and Reporting

1. Introduction

Internet-based research presents research companies with a different cost model. Instead of the variable-based cost model of telephone interviewing, the technology associated with Internet-based research employs a fixed-cost model which requires higher level of investment that must be amortized across a larger number of interviews. Market research companies that pursue the building and management of Internet data collection infrastructure and systems must adopt mainstream Information Technology processes and methodologies and higher levels of spending, or failure is certain. Some of the technical challenges that must be addressed include:

- Hardware (internal and external network, servers, desktops) procurement, installation and management
- Internet bandwidth (ISP) procurement, installation and management
- Software requirements, development, quality assurance (including usability, load and stress testing), and documentation
- Database management
- Information security
- Disaster recovery and business contingency planning

- Technical project management for complex hardware and software projects
- Help Desk implementation and management
- Management, development and retention of technical staff
- Management of external consultants
- Vendor management

This paper addresses the issues facing market research agencies who are managing or planning to manage Internet research technology infrastructure. However, the cost and complexity associated with Internet research technology infrastructure will cause many companies to elect to address this capability via outsourcing arrangements. The issues addressed herein are also applicable to companies that choose outsourcing arrangements and this paper can be used as guidance or to develop a "check list" when working with or selecting an outsourcing partner.

Note that data collection is only one component of an intricate, interdependent Internet research system. If you have system for data collection, you still have to deal with myriad of other issues such as database creation, maintenance information security, and results distribution. You need the software, the expertise, the hardware and the people to address these issues.

2. Sampling and Panel Management

There are many offerings in the nature of systems available for Internet sampling and panel management. Most of these, however, are not designed with the needs of the market researcher in mind. Meeting the needs of sophisticated market research (a current trend that the authors have observed is to move complicated CATI surveys to the Internet) requires sophisticated sampling and panel management tools. These are generally custom systems that are built by research companies or outsourcing partners.

After sample is selected and processed, it needs to be e-mailed and tracked. This means tracking "bounce back" e-mails (e-mails that "bounce back" to the sender due to a problem such as an invalid e-mail address), determining the cause of the "bounce back" and removing invalid e-mail addresses from the sample or panel. With the coming regulations regarding unsolicited commercial email ("Spam"), it's critical that research providers follow appropriate confirmed opt-in and list removal rules (which may require yet another server solution or outsourced arrangement). One organization that publishes standards for online research is the Council of American Survey Research Organizations (CASRO). For more information, please see the CASRO web site at www.casro.org.

3. Load Balancing, Authentication, Survey Re-entry

There are three areas that need to be addressed when a respondent comes to an Internet research site to complete a web survey. These are load balancing, authentication, and survey re-entry.

- Load balancing (dividing work between multiple servers in order to provide the appropriate performance and redundancy): In order to achieve the required scale (Internet samples are larger than telephone samples and large numbers of respondents hit the servers simultaneously) and redundancy, Internet research requires the deployment of multiple servers for collection of interviews. This requires a system to appropriately balance the load of surveys across multiple servers. As the size of the server "farm" grows, and the number of projects performed simultaneously increases, load balancing takes on greater significance. The overloading of a

single server can result in unpleasant respondent experiences at best, and system crashing, with resulting data loss, at worst.

- Authentication (the process of verifying a user/respondent's identity): This entails verifying that the proper respondent takes the proper survey, typically requiring a password.
- Survey re-entry: This involves making sure that a suspended or halted survey is re-started at the appropriate place and that prior collected data is available for subsequent survey logic.

4. Scalability and Reliability

With the move to Internet research, samples typically are larger and more compact in time than CATI surveys. This, combined with the fact that a large research organization fields many surveys at a time (at Harris Interactive, this number is typically in excess of 100) presents a need to scale to large peak volumes.

The challenge is to build an infrastructure that can scale to large numbers and do so reliably. Suppose you are averaging 10,000 on line interviews per week, and you get a contract requiring 20,000 per week. Will the current infrastructure handle the increase load? If not, how much is it going to cost to increase your infrastructure to handle the new load? The answer to this question requires that you understand:

- The timing of the load (all at once or evenly spread over some time).
- The number of simultaneous interviews the system can handle before response time degrades to an unacceptable level (or the system crashes).
- The bottlenecks to increasing load.
- The cost of increasing the performance level of those bottlenecks.
- The capacity of your Internet bandwidth to handle the increased traffic.

In order to answer many of these questions, it is necessary to perform a stress test of your production infrastructure. A proper stress test simulates hundreds or thousands of Internet respondents "attacking" you system at the same time, and monitors all aspects of the system. Increased load should be put on the system until a failure point is reached, at which time the accumulated data can be analysed to understand the cause of the failure, and the capacity of each node.

It is said that in a well-tuned system, all components will fail simultaneously. As that probably won't happen, the failed node will be the bottleneck to increased performance. Once this bottleneck is identified, and more resources put against it, the stress test should be repeated until either:

- The system proves capable of handling the desired load, or
- The next bottleneck is identified and upgraded.

The implementation of such a test requires specialized hardware, software, and software quality assurance expertise that only the largest IT Engineering staffs are equipped to provide. Smaller firms may wish to outsource such tests to specialty firms such as Mercury Interactive (http://www-heva.mercuryinteractive.com).

5. Security

Security is a critical area for any company engaging in any kind of commerce on the Internet. For research companies, security falls into the following areas:

- Security of infrastructure from hackers and denial of service attacks

- Security of client information, confidential respondent information
- Protection of web objects served as part of the interview from respondent capture

Security is more than firewalls (a firewall is a combination of hardware and software located at the gateway or entry point to a network that protects resources on that network). It requires a strong security policy that is enforced and followed by all employees. It requires the resources to log break-in attempts, and the skilled personnel to follow-up on those attempts to record any that are successful. Only by knowing about successful penetrations can steps be taken to thwart them in the future. Diligent security requires monitoring of software manufacturer patches, and applying them in a timely matter to production systems.

Security measures are not complete unless they are tested. Ideally, the IT Director should hire a security company to stress his defences by attempting to hack into the system.

6. Survey Functionality

One of the recent challenges faced by telephone research is the continual decline in cooperation rates. This is due to a variety of factors including the negative impact of telemarketing, the busier lifestyle people live today (multiple earner households result in family time that is scarcer and less willing to participate in telephone research), the abuse of respondents by many research companies (long complex CATI interviews), etc. This, combined with the move to permission based marketing and market research, makes Internet research more and more attractive. One trend the authors have observed (as mentioned above) is the move of complex CATI studies to the Internet. So, a key to success here is that the Internet research software used has to have at least all the functionality that clients are accustomed to with CATI software. This means Internet research software with a full set of functionality. Required features include:

- All the questionnaire features found in mainstream CATI software
- Easy incorporation of Internet capabilities like graphics, multimedia, etc.
- Quota management
- Ability to easily implement different templates
- Ability to support any language
- Ability to include advanced research techniques

Additionally, the Internet is a visual medium, meaning that there is the potential to put more than CATI-like functionality in front of the respondent. The use of plug-in tools, audio, video, and images means that more data can be gathered, in less time, with less respondent burden, than with other interviewing modes. Your web interviewing software must include the capability to do all of these, and make it easy to construct and deploy rich survey media.

7. Deployment and Re-deployment

Once a questionnaire has been programmed, it is ready for testing. Good survey software has the ability to deploy the questionnaire and any ancillary material (templates, graphics, video files, etc.) to the server(s) so it can be tested by researchers and clients. Testing needs to include the email transmission (in multiple languages, potentially), authentication mechanism (verification of user/respondent identity), URL accessibility (link to the Internet survey) both from within the agency

firewall and without, and the questionnaire logic itself. Quota mechanisms which may change the questionnaire logic should also be tested.

When changes are necessary, the system needs a rapid way to re-deploy the revised questionnaire, so testing can be resumed.

Once testing is complete, and the survey has been opened to the respondents, the survey is still subject to change. These changes need to be incorporated into the questionnaire, data model (conceptual model that represents survey data, metadata, etc.), quota controls, graphics files, and anything else having to do with the execution of the project, and it must all be done while the survey is in the field, without upsetting the respondents already in process.

8. Management/Quota Management

Critical to the Internet research process is the ability to apply the management of the data collection process that researchers are accustomed to with CATI. This means:

- Survey management reports tools (topline, disposition reports)
- Quota status reporting and management tools
- The ability to add, delete, and change quotas on the fly, and in real time.
- The ability to have very large numbers of quotas (tens of thousands, in some cases)

A key here is that all these tools should be web based and must empower the researcher and client. In order to satisfy the tighter timeframe associated with Internet research, researchers and clients must be able to use these tools at their convenience and the convenience of the research.

9. Aggregation, Data Weighting and Reporting

As stated above, the tighter time frames associated with Internet research require web-based tools that are available to researchers and other survey professionals (such as statisticians). Requirements:

- Data must be aggregated to a centralized data store (possibly a relational database) in near real time (typically 15 minutes after survey completion)
- Weighting tools must be able to access this centralized data store and permit high-speed generation and application of weights (typically within hours of completion of survey field period).
- Open ended coding and response recodes must be performed contemporaneously with data collection to meet the new expectations for data delivery.
- The centralized data store with weights applied must be available to web-based (online) reporting tools. Weighting problems have become more sophisticated with the need for behavioural variables as well as demographic population targets.

These online reporting tools fall into two basic categories:

- Real time reporting and survey management. These reporting systems typically provide frequency, crosstabulation, survey disposition, and verbatim (open end) reporting.
- Customized web reporting. These systems include customized views driven by client requirements and are organized around client organizational charts or bureaucracy.

10. Conclusions

There is much more to establishing the ability to gather web based questionnaire responses than shrink-wrapped survey engine software. We have evolved passed “forms on the screen” surveys. A complete system is one that ensures a quality experience for the respondent, and leaves him wanting more. It takes many system components to accomplish this, many of which are not available off the shelf. This makes the business of web based data collection very capital intensive, as well as one which requires a highly skilled technical staff to maintain.

About the Authors

Leonard Bayer, CCP is Executive Vice President, Chief Technology Officer, and Chief Scientist of Harris Interactive Inc. At Harris Interactive, Len is responsible for Information Technology strategy, including overall direction for technology research and development. He also guides researchers in the areas of marketing science, research methodology, and statistics. Most of Len’s time is spent on Internet related research, and in front of customers, partners, and researchers evangelising Harris Interactive technology. Len holds a B.S. in Astrophysics, a B.A. in Mathematics, and an M.A. in Mathematical Statistics from the University of Rochester. He has finished all the requirements for a Ph.D. in Mathematical Statistics except for the completion of the dissertation. He is also a Certified Computer Professional (CCP). He can be contacted at Harris Interactive, 135 Corporate Woods, Rochester, NY, 14623, 585-214-7417, fax, 585-273-0500, e-mail lrbayer@harrisinteractive.com.

Peter Milla is Executive Vice President and Chief Information Officer at Harris Interactive Inc.. He has 20 years of experience in a wide range of survey research and data processing activities, applying skills in both computer technology and survey research methods. Peter has extensive experience with all CASIC (Computer Assisted Survey Information Collection) technologies, including Internet-based Survey Research. He also has expertise in LAN/WAN design, implementation, and operation as well as large-scale computer software installation and training. Peter holds a B.A. and M.A. from Queens College of the City University of New York. Peter is currently a member of the Board of Directors of the Council of American Survey Research Organizations (CASRO). He also serves as chair of CASRO’s Technology Committee.. He can be contacted at Harris Interactive, 111 Fifth Avenue, New York, 10003; tel. 212-539-9705; fax 585-273-0511; e-mail pmilla@harrisinteractive.com

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Remote Access of Microdata

Barry Schouten

Abstract

Statistical agencies manage huge amounts of microdata. The main task of these agencies is to provide a variety of users with general information about for instance the population and the economy. However, in some cases users request additional, more specific information. Many agencies have therefore set up facilities that enable selected users to obtain tailor-made statistical information.

A remote access system is an example of such a facility where users can submit queries for statistical information from their own computer. These queries are handled by the statistical agency and the generated, possibly confidentialised, output is returned to the user. This way the agency still keeps control over its own data while the user does not need to make frequent visits to the agency.

At Statistics Netherlands and at other statistical institutes recently the need for a similar system has been expressed. In the presentation, we will briefly discuss the characteristics, limitations and desired properties of a remote access system.

Keywords

Remote access, Microdata, Data dissemination, Confidentiality

1. Introduction

To achieve transparency of governmental policies and to facilitate research, statistical agencies publish information that is generally accessible. However, governmental organisations, universities and other research institutes often need more specific information. Many agencies have acknowledged this need and have implemented facilities to handle these specific, ad hoc requests.

One such facility is a remote access system where researchers can submit statistical queries via a web connection from a remote location, usually their own place of work. The microdata remain physically at the office of the statistical agency. The agency executes the queries, evaluates the disclosure of individual information, if necessary controls the output and returns the confidentialised output to the users.

The protection of the confidentiality of microdata and the dissemination of statistical information are two conflicting objectives. Statistical agencies have to take a stand based on the legislations and based on their own ethics. For a discussion see for instance Duncan and Pearson (1991) and Schouten and Cigrang (2003).

Due to their double role, most agencies have been rather conservative when it comes to granting access to their microdata. In the literature over the last 15 years relatively little attention has been paid, therefore, to remote access of statistical microdata. Exceptions are, for instance, Cigrang and Rainwater (1990), Keller-McNulty and Unger (1998), NISS (1998) and Blakemore (2001).

Recently at several statistical institutes the need for remote access systems has been expressed. To illustrate this we refer to the papers by Andersen (2003), Coder and Cigrang (2003), Desai (2003), Franconi and Merola (2003) and Shlomo (2003) presented at the joint UN/ECE and Eurostat work session on statistical data confidentiality in Luxembourg. The papers can be downloaded from the UN/ECE website, see http://www.unece.org/stats/documents/2003.04.confidentiality.htm.

Two remote access systems are worth mentioning here, namely the Lissy system of the Luxembourg Income Study (LIS) and the Digital Government Program in the United States set up by the National Institute of Statistical Science (NISS). See, respectively, http://www.lisproject.org and http://www.niss.org/dg.

Also, recently, at Statistics Netherlands a pilot was started which ultimately has the goal to construct a remote access system for governmental organisations like ministries and planning offices. The remote access system should give access to various household surveys linked to register data for research purposes.

However, before and during implementation of a remote access system there are a number of important questions that need to be addressed and that statistical agencies are faced with. What stand should we take between confidentiality and dissemination? What statistical queries do we allow? What are the implications for confidentiality evaluation and control? In what way should we rely on contracts with the researchers? Can we shift part of the disclosure evaluation to a point in time after the release of the output? What about the security of the database and the web connection?

In the presentation we will focus on a number of these questions, discuss properties and limitations of remote access systems and comment on some of the preliminary outcomes and experiences at Statistics Netherlands.

Basically, there are three main features that play a central role in the implementation of a remote access system, see Blakemore (2001) and Schouten and Cigrang (2003). In the next section we will elaborate briefly on these features. In section 3 we will summarize the tentative findings of the pilot at Statistics Netherlands and make some recommendations. A discussion follows in the final section.

2. Confidentiality, feasibility and user-friendliness

The basic features that play a role are confidentiality, user-friendliness and feasibility. In the literature user-friendliness is also referred to as utility.

Confidentiality concerns the protection of data on individual persons, households, companies and other institutions.

First, one needs to decide what microdata to make accessible in the remote access system. Clearly, this decision depends on the needs of the users, the relevance of the microdata for research or policy making, the legislation, the quality of the data and the sensitivity of the information contained in the microdata.

Next, it should be decided whether the statistical database is static or dynamic, i.e. whether it contains longitudinal data and whether it will be updated in time. This has consequences for the disclosure control.

Legislation needs then to be interpreted so that disclosure can be defined. Especially, when remote access systems are interactive and disclosure evaluation is (partially) automatic, clear definitions are essential and inevitable.

When clear definitions of disclosure are available, then the next step is to decide how confidentiality of the data can be assured. This means disclosure must be evaluated and controlled. Disclosure evaluation and disclosure control cannot be viewed separately. The types of output that are not allowed determine the measures that must be taken to change the output.

Evidently, disclosure evaluation and control can in many cases not be constructed so that the remote access system always guarantees the confidentiality of the data. Different users may have different levels of prior knowledge or may combine output with output from previous queries. Therefore, besides the removal of all direct identifiers, the microdata can be locally suppressed, perturbed, masked or recoded. Disclosure is then partially prevented by adapting the microdata or by increasing the uncertainty in the generated output.

As a last step the statistical agency can demand that users of the remote access system sign a contract in which they agree on a number of rules and in which the consequences of violations are described. Setting up a contract does not only serve as a back-up in case of misuse of the system, but will also deter potential misuse because of the consequences.

Another main feature of a remote access system is its user-friendliness or its utility to users. First it must be decided which users are granted access. Next, relevant aspects of the remote access system to the user are speed, documentation and the presence of a helpdesk.

The service time of the system depends on the extent to which the system is automated, i.e. are queries evaluated and executed manually. Furthermore, the choice for an on-line (webpage) or off-line system (via e-mail), and the number of users also influence the speed with which queries are replied. Finally, and maybe most importantly the speed is determined by the evaluation and control of disclosure.

To both the administrators and the users it is vital that the system is well documented, so that correct queries can be formulated and that the output is clear and interpretable.

Essential to the architecture of the remote access system is the feasibility of its implementation. The most important aspect of the feasibility is the security that can be attained. Although, queries and output are evaluated for disclosure, the statistical database itself and the data transfer between the user and the agency may be the subject of hackers.

The capacity needed for the maintenance and administration of the system, and to keep up a certain level of service also play a role when it comes to feasibility. Furthermore, the choice for automatic or manual handling and evaluation of queries determines the labour intensiveness of the remote access facility. These aspects are all directly related to budgetary constraints.

Ideally, the three features confidentiality, user-friendliness and feasibility should all be optimal at the same time. However, these features are clearly conflicting. When setting up a remote access facility one should balance confidentiality and user-friendliness under feasibility constraints. This means the statistical agency has to decide what disclosure evaluation and control actions to employ before output is released, what actions to employ afterwards, and in what way they rely on the adaptation of the

underlying microdata and the use of contracts. Furthermore, the feasibility constraints may be gradually relaxed in time based on the evolving experiences with the remote access system concerning the disclosure evaluation, the number of users and the speed of the system.

3. The pilot at Statistics Netherlands

The pilot at Statistics Netherlands was initiated in the summer of 2002. The Dutch Ministry of Social Affairs and Employment participates in this pilot by submitting queries in SPSS by e-mail. With these queries a number of employees of the ministry can perform statistical analyses on a large microdata file containing information on social allowances. The file contains over a million records over the period 1997-2000. Only the direct identifiers were removed from the microdata.

Statistics Netherlands has supplied the researchers with documentation on all variables involved. Also a test file was constructed that contains a 1% sample of the records in the microdata. The purpose of the test file is twofold. First it serves to test whether the syntax of a query is correct. Secondly, the tes file offers the users a first glance on the variables. For this reason only the demographic variables in the test file are made meaningless by data swapping.

The submitted queries are run on a computer and are returned to the user after disclosure evaluation. The output is returned to a prescribed e-mail address and supplied with a password. Shortly after receiving the query and after forwarding the output, receipt of the e-mail is acknowledged by phone calls to the researcher. The process of receiving, running and returning queries is completely structured by a checklist. The checklist contains all the actions to be taken and the employees responsible for each action.

Clearly, queries that print individual records or change the microdata in any way are not allowed. In principle every statistical analysis that is supported by SPSS can be submitted. Researchers are also allowed to define new variables and to select subpopulations. However, the users are informed beforehand that some commands may lead to a rejection of the query. For instance, the use of several consecutive selections may result in subpopulations that are too small.

We intentionally started the pilot at a low level when it comes to the number of users and the automation of query handling. The disclosure evaluation and control is completely manual. Every query and its corresponding output is logged and a summary report is made of the evaluation of the output and subsequent control measures. One of the main reasons for starting the pilot at a low level was that we wish to set up guidelines for the evaluation of the output and to construct simple checks. These simple checks may be used to ultimately make an output filter and should be more restrictive than is strictly necessary, so that output passing a future output filter can never compromise confidentiality. Output that is filtered, then potentially discloses information on an individual level and is evaluated manually.

Furthermore, since it is our intention to build a system that allows detailed and flexible research, it was decided to filter especially at the output level. Clearly, this is only possible whenever there is a relatively small number of users. In case the pilot is extended to a large number of researchers, it is inevitable that part of the review is automated or that part of the evaluation is shifted to the input level.

The manual review in the pilot was very useful and provided us with a lot of insight in the queries that are submitted. Here we present some of our preliminary findings:

- It is essential that the review of output is structured in order to make disclosure evaluation consistent and less time-consuming. The following measures may be taken:
  - Rank every variable according to the extent that it identifies individuals, i.e. whether it represents a visible feature or a feature that is traceable in generally accessible databases. Consequently, a taxonomy of variables can be made based on this ranking.
  - Identify variables that are extra sensitive, so that extra care can be taken whenever these variables appear in the output.
  - In case the microdata come from an administrative register or from a subpopulation, decide whether the selection is identifiable. If the selection variable is not a visible or traceable feature, then the evaluation can be less restrictive.
  - Decide for the most recurring types of analyses what output has the potential to disclose individuals and what output is never allowed.
- The quality of the microdata, sufficient documentation, and technical and scientific support are absolute prerequisites for a user-friendly system. In case the microdata contain many administrative errors researchers will in general need more queries to produce reliable results. It seems imperative to invest in these aspects beforehand.
- Automation of output review can only be implemented in case flow charts are made that give simple and sufficient checks.
- For the evaluation of disclosure it is essential that the users supply their queries with sufficient remarks on the objectives and insert comments before the most important statistical commands. New variables must be provided with clear names and value labels. Importantly, the addition of explanatory comments also serves to evaluate whether the analysis complies with the objective of the research as stated by the user.
- In case some commands recur often, it is efficient to allow researchers to save their own adapted microdata files. These files should then be documented and evaluated before further use. Also other microdata from other sources may be linked.
- Logfile summary reports may best be represented in a fixed format containing for instance the name of the microdata file, the variables used, the variables used that are visible or traceable, the variables used that are sensitive, new variables and their classifications, selections, the relation to previous queries and comments on the evaluation and control of disclosure.
- The use of a checklist with the persons responsible for each action aids the speed with which queries are handled.

Many statistical agencies have also set up on-site facilities. These facilities are usually located within the offices of the agency. Researchers can perform analyses on stand-alone computers containing microdata files and only the output that they wish to take with them is evaluated. An important question is whether remote access should be offered as an alternative for on-site access.

For some years Statistics Netherlands also has on-site facilities at two locations in the Netherlands. In order to be granted access a researcher needs to write a research proposal and must sign a contract. The pilot allows a comparison with these on-site facilities.

The users of the on-site facilities are not limited in the statistical queries that they run on the microdata. Only the direct identifiers are removed from the microdata beforehand. However, the output that users wish to take with them must satisfy the same rules as the output that is generated by remotely submitted queries.

Clearly, the primary reason for setting up a remote access facility is its user-friendliness. A major drawback to the users is that some output will be filtered. For instance, a scatterplot of the residuals of a linear regression versus the independent variables will usually not pass the manual review. There are ways to overcome some of these restrictions. Reiter (2003) suggests synthetic residuals in order to

give an idea about the quality of the model fit. See also Reiter (2002). However, remote researchers will always encounter some limitations.

To the data provider the advantage of remote queries is that no intermediate queries and output are missed. Output from on-site research has the tendency to be rather compact and more difficult to evaluate since some intermediate steps are missing. Against this one can state that remote output has the tendency to be more lengthy and may also contain output that will turn out to be uninteresting to the user.

The pilot shows that remote access is especially useful for queries that are relatively simple and for queries that are supplementary to more thorough statistical queries.

4. Discussion

In section 2 we gave a description of the main features of a remote access system. We refer to Schouten and Cigrang (2003) for a more detailed discussion.

Contracts and adaptation of the original statistical database can be viewed as measures to reduce the risk that remains after implementing a set of disclosure evaluation techniques. It is important, therefore, to decide in what respect we may rely on the trustworthiness of the users of the remote access system. Especially, when it comes to specific users like governmental organisations, universities and other research institutes, one may argue whether it is necessary to filter all output that only by very sophisticated methods leads to potential disclosure. Such users make a lot of effort when intentionally searching for individual information. One other argument that is often put forward is that in the past only very few violations have been reported. However, there must be a reasonable level of data confidentiality in order to conclude that violations do not occur and to show that disclosure is taken very seriously by the agency.

In the pilot we found that it is essential that queries are well documented by the users. The documentation of variables and queries is regarded as burdensome by the researchers, but it aids the speed with which the generated output is reviewed and it also signals that we really consider the confidentiality of the output.

Another benefit of documenting queries is that the documentation can be used to evaluate whether the queries comply roughly with the research objective. Queries that do not comply might indicate that the user has other intentions.

In general, we believe that remote access systems are valuable alternatives for other dissemination methods. The pilot confirms us in our belief. However, one must fix an acceptable level of disclosure evaluation, and as for most data dissemination a small risk of disclosure must be accepted. In order to balance user-friendliness and confidentiality part of the disclosure evaluation must be performed after the output has been released. The risk that originates form this partial shift of evaluation must be covered by the contract.

Challenging research objectives for the future are: the construction of logfiles and summaries of logfiles that preserve the information needed to evaluate disclosure, the balancing of confidentiality and user-friendliness for sophisticated statistical analyses, and the automation of disclosure evaluation in the form of input and output filters.

References

Andersen, O. (2003), From On-site to Remote Data Access – The Revolution of the Danish System for Access to Microdata, Paper presented at the joint UN/ECE and Eurostat work session on Statistical Data Confidentiality, Luxembourg, April 7–9 2003.

Blakemore, M. (2001), The Potentials and Perils of Remote Access, in *Confidentiality, Disclosure, and data Access: Theory and Practical Application for Statistical Agencies* (eds. P. Doyle, J.I. Lane, J.J.M. Theeuwes, L.M. Zayatz), Amsterdam, Elsevier.

Cigrang, M., Rainwater, L. (1990), Balancing Data Access and Data Protection: The Luxembourg Income Study experience, Proceedings of the American Statistical Association, Section on Statistical Computing, pp. 1-4.

Coder, J., Cigrang, M. (2003), Lissy Remote Access System, Paper presented at the joint UN/ECE and Eurostat work session on Statistical Data Confidentiality, Luxembourg, April 7–9 2003.

Desai, T. (2003), Providing Remote Access to Data: the Academic Perspective, Paper presented at the joint UN/ECE and Eurostat work session on Statistical Data Confidentiality, Luxembourg, April 7–9 2003.

Dobra, A., Karr, A.F., Sanil, A.P., Fienberg, S.E. (2002), Software Systems for Tabular Data Releases, International Journal of Uncertainty, Fuzziness and Knowledge Based Systems 10 (5): 529-545.

Duncan, G.T., Pearson, R.B. (1991), Enhancing Access to Microdata while Protecting Confidentiality: Prospects for the Future, Statistical Science 6: 219-239.

Franconi, L., Merola, G. (2003), Implementing Statistical Disclosure for Aggregated Data Released via Remote Access, Paper presented at the joint UN/ECE and Eurostat work session on Statistical Data Confidentiality, Luxembourg, April 7–9 2003.

Hawala, S. (2001), American FactFinder: U.S. Bureau of the Census Works towards Meeting the Needs of Users while Protecting Confidentiality, UNECE Statistical Division, Working Session on Statistical Data Confidentiality, Skopje.

Keller-McNulty, S., Unger, E.A. (1998), A Database System Prototype for Remote Access to Information based on Confidential Data, Journal of Official Statistics 14: 347-361.

National Institute of Statistical Sciences (1998), A Web-based Query System for Disclosure-limited Statistical Analysis of Confidential Data, Research proposal of Digital Government project available through http://www.niss.org/dg.

Reiter, J. (2003), Synthetic Diagnostics for Remote Access Regression Servers, To appear in Statistics and Computing 4.

Reiter, J. (2002), Satisfying Disclosure Restrictions with Synthetic Data Sets, Journal of Official Statistics 18, no. 4, 531–543.

Schouten, J.G., Cigrang, M. (2003), Remote Access Systems for Statistical Analysis of Microdata, to appear in Statistics and Computing 4.

Shlomo, N. (2003), Accessing Microdata via the Internet, Paper presented at the joint UN/ECE and Eurostat work session on Statistical Data Confidentiality, Luxembourg, April 7–9 2003.

About the Author

The author is a statistical researcher at the Methods and Informatics Department of Statistics Netherlands. The views in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Generalized Linear Model Diagnostics for Remote Access Servers

Jerome P. Reiter & Christine N. Kohnen

Abstract

To protect public-use microdata, one approach is to allow users to submit analyses to a remote computer that reports back basic output from the fitted model, such as coefficients and standard errors. To be most useful, the server also should provide some way for users to check the fit of their models, without disclosing sensitive data. This paper discusses remote server model diagnostics for commonly used generalized linear models, including linear, logistic, and multinomial regression models. The diagnostics are illustrated with simulations.

Keywords

Confidentiality, Diagnostics, Disclosure, Generalized linear model, Remote access

1. Introduction

Statistical agencies that release public use microdata seek to provide users with detailed data that can support a wide range of analyses, while guarding the confidentiality of survey respondents' data. Protecting confidentiality usually has priority, so that the collected data are often altered before public release. Typical methods for altering data include recoding, swapping, or adding noise to the original data (Willenborg and de Waal, 2001). However, these methods can compromise estimation by distorting relationships among variables in the data set. Furthermore, they complicate analyses for users. For example, to analyze properly data that include additive random noise, users should apply measurement error models (Fuller, 1993).

To circumnavigate these problems, agencies can require users to submit analyses to a remote computer that reports back basic output from the fitted model, such as coefficients and standard errors, but never allows the users access to the original data. The remote access approach has advantages over methods of disclosure control that swap, recode, or add noise to the original data: analyses are based on the original data, and standard analyses can be used. Additionally, confidentiality may be protected more effectively since no actual microdata are released. See Keller-McNulty and Unger (1998), Duncan and Mukherjee (2000), and Schouten and Cigrang (2003) for discussions of these aspects of remote access approaches.

The remote server also should provide some way for users to check the fit of their models. Unfortunately, releasing the usual diagnostic statistics can disclose values. For example, when actual residuals and fitted values (values of the dependent variable on the estimated regression line) are released for a submitted linear regression model, the user can obtain the values of the dependent

variable by simply adding the residuals to the fitted values. A way around this problem was proposed by Reiter (2003): remote servers can provide synthetic, i.e. simulated, diagnostics that mimic real-data diagnostics. For example, in linear regression models with Gaussian errors, agencies can release synthetic values of dependent and independent variables, residuals, and fitted values. Users then can treat these synthetic values like ordinary diagnostic quantities, such as examining scatter plots of the synthetic residuals versus the synthetic independent variables or versus the synthetic fitted values.

In this paper, we extend this approach to remote server diagnostics for logistic and multinomial regression models. The paper is organized as follows. Section 2 reviews the remote server diagnostics for linear regressions developed by Reiter (2003). Sections 3 and 4 discuss remote server diagnostics for logistic and multinomial regression models. Section 5 concludes with a brief discussion of other issues in remote servers.

2. Remote Server Diagnostics for Linear Regressions

For linear regressions, model diagnostics frequently involve the analysis of residuals. A remote server, however, must protect confidentiality and so generally cannot release the exact values of the residuals. Fortunately, for some types of diagnostics in linear regression--such as plots of residuals versus independent variables--the exact values of the residuals and independent variables are not needed; rather, the relationships among the residuals and independent variables are examined for patterns in hopes of identifying model misspecifications. Thus, for remote server diagnostics, it may be adequate to mimic patterns in the real-data diagnostics rather than release real-data values.

To generate the synthetic diagnostics, the remote server administrator first simulates synthetic values of each variable in the data set. The details are in Reiter (2003), but we outline the process here. Let $\mathbf{x}_p$ for $p = 1, \ldots, d$, denote a variable in the data set, and let $\mathbf{x}_p^s$ denote the synthetic version of $\mathbf{x}_p$. Each $\mathbf{x}_p^s$ can be generated by drawing from an estimated marginal density of $\mathbf{x}_p$. The same $\mathbf{x}_p^s$ are provided to all users of the remote server, regardless of the users' queries.

When a regression is submitted to the server, the server generates and provides synthetic, standardized residuals for use in plots of synthetic residuals versus synthetic independent variables. Let *k* index one of the synthetic values in $\mathbf{x}_p^s$. Let t_{kp}^s be the synthetic, standardized residual attached to synthetic value *k* of variable *p* in the fitted regression. We seek to generate the t_{kp}^s so that the relationship between $\mathbf{t}_p^s = \{t_{kp}^s\}$ and synthetic $\mathbf{x}_p^s$ is similar to the relationship between the real-data standardized residuals, $\mathbf{t}$, and the real-data $\mathbf{x}_p$. To this end, we determine each t_{kp}^s as follows:

$$t_{kp}^s = b_{kp} + v_{kp} + n_{kp}.$$

The b_{kp} places t_{kp}^s on a curve consistent with the general relationship between $\mathbf{t}$ and $\mathbf{x}_p$. The $\mathbf{v}_{kp}$ moves the synthetic residual off that curve in a way that is consistent with the variation in the $\mathbf{t}$ near $\mathbf{x}_p = k$. The n_{kp} is noise added to reduce the risk of disclosing the values of the real-data residuals. These three pieces are described in detail below.

To determine b_{kp} when *p* indexes a continuous independent variable, a smooth curve is fitted to the relationship between $\mathbf{t}$ and $\mathbf{x}_p$ using a generalized additive model (Hastie and Tibshirani, 1990). The

b_{kp} is equal to the value of this curve at *k*. Administrators can use any specification of the generalized additive model that describes the relationship between $\mathbf{t}$ and $\mathbf{x}_p$. When *p* indexes a categorical variable, $b_{kp} = 0$ because a smooth curve between $\mathbf{t}$ and $\mathbf{x}_p$ is not needed; all values in $\mathbf{x}_p^s$ are in $\mathbf{x}_p$.

To determine each v_{kp}, the unit *j* is found such that $j = \arg\min_i \left| k - x_{ip} \right|$; this is the unit *j* whose value in the real-data $\mathbf{x}_p$ is closest to the synthetic value *k*. When more than one unit satisfies the arg-min condition, unit *j* is obtained by sampling randomly from the qualifying units. When *p* indexes a continuous independent variable, the $v_{kp} = t_j - b_{jp}$, where b_{jp} is the value at x_{ip} on the curve obtained from the generalized additive model. When *p* indexes a categorical independent variable, the $v_{kp} = t_j$. Effectively, this randomly selects a standardized residual from the units with $x_{ip} = k$.

Each n_{kp} is drawn from independent $N(0, \tau)$, where τ is specified by the administrator of the remote server. Different values of τ can be used for different regressions. However, a single τ is used for all synthetic residuals from the same regression, so as not to introduce artificially non-constant variance in the synthetic residuals. All queries that use the same dependent variable should use a common random seed to generate the n_{kp}. This prevents users from refining any guesses about a real-data t_i by averaging synthetic residuals from repeated calls to the same or similar regressions. Reiter (2003) argues that setting $\tau = 1$ generally should provide reasonable protection for units fitting close to the regression line, since prediction intervals for dependent variables based on the synthetic residuals should have the same width as those based on the root mean squared error of the regression. Units with large t_{kp}^s may need to be top-coded, i.e. released as "$t_{kp}^s \geq C$".

Because these diagnostics are synthetic, they may miss some model inadequacies that can be revealed using real-data diagnostics. That said, Reiter (2003) shows via simulation studies that these synthetic diagnostics improve data utility relative to releasing only the basic output from the fitted model.

3. Remote Server Diagnostics for Logistic Regressions

Releasing real-data, unit-specific logistic regression diagnostic measures, such as residuals or case influence statistics, can result in disclosures. All units with outcome equal to one must have positive residuals, and all units with outcome equal to zero must have negative residuals. Hence, when the user is able to determine the sign of a unit's residual, he or she knows the value of that unit's outcome. Similar problems exist for popular case influence statistics, such as the change in the total deviance or change in the chi-squared test statistic after deleting each observation from the data set (Pregibon, 1981; Hosmer and Lemeshow, 2000). Any unit with a large case influence statistic does not fit the pattern of the model, so that its value of the outcome must be as far as possible from its predicted probability. In other words, units with large case influence statistics and small predicted probabilities have outcomes equal to one, and units with large case influence statistics and large predicted probabilities have outcomes equal to zero.

Adding noise to the residuals or case influence statistics generally will not transform them into simultaneously safe and useful diagnostics. For residuals, the noise must be large enough to cause users to be unsure of the signs of the actual residuals; this practically eliminates the diagnostic utility

of the synthetic residuals. For case influence statistics, adding random noise leaves similar problems: when the noise is small, the user still can associate large values with the units' outcomes, and when the noise is large the diagnostics have limited diagnostic utility. Hence, we do not pursue case-specific, remote server diagnostic measures for logistic regressions.

We propose instead to release grouped diagnostics. This proceeds in four steps, all performed separately for each $\mathbf{x}_p$: (i) partition $\mathbf{x}_p$ into m_p categories; (ii) calculate the percentages of ones in each of the m_p categories, and add a small amount of noise to these percentages; (iii) calculate the averages of the predicted probabilities from the model for each of the m_p categories; and, (iv) release the medians of the $\mathbf{x}_p$, the perturbed percentages of ones, and the averaged predicted probabilities. When the predicted probabilities are substantially different from the perturbed percentages, the model needs to be adjusted to fit better in the problematic regions of $\mathbf{x}_p$.

Random noise is added to the observed percentages of ones to prevent disclosures for categories of $\mathbf{x}_p$ in which the outcomes are almost all zeros or all ones. We perturb the number of ones in each category of $\mathbf{x}_p$ by randomly adding a uniform draw from (-2, -1, 1, 2). When the resulting, perturbed percentage of ones in any category is less than zero or greater than one, we redraw until the perturbed percentage is in fact between 0% and 100%. Adding or subtracting a maximum of two units ensures that any released percentage could have been generated from at least two values of the real-data percentages, which helps protect confidentiality. Excluding zero from the set of possible random draws ensures that no exact percentages are released. The same random seed should be used to generate the random noise for all logistic regressions involving the same outcome variable, for reasons discussed in Section 2.

For categorical $\mathbf{x}_p$, the m_p categories can be the levels of $\mathbf{x}_p$, provided there are sufficient numbers of units in each level so as to prevent disclosures. For continuous $\mathbf{x}_p$, we recommend forming categories with at least 100 units with similar values of $\mathbf{x}_p$, since adding or subtracting noise of at most 2% should not greatly distort the relationships between the real-data percentages and predicted probabilities for most values of these quantities.

To illustrate the utility of these diagnostics, we simulate two scenarios using different specifications for the independent variables in the logistic regressions. For each simulation, we generate $n = 10{,}000$ observations. The independent variables, x_1 and x_2, are randomly drawn from $N(0, 2^2)$. To generate the dichotomous dependent variables, we draw from Bernoulli distributions with probabilities, $\exp(g(x_1, x_2))/(1+\exp(g(x_1, x_2)))$, where the $g(x_1, x_2)$ are as shown in Table 1.

Table 1: Scenarios for logistic regression simulations

Scenario	Function of predictors
Non-linearity	$g(x_1, x_2) = -3 + 4.3x_1 + 1.5x_1^2$
Interaction	$g(x_1, x_2) = 1 + 2.2x_1 - 4.6x_2 + x_1 x_2$

We generate grouped diagnostics using the methods outlined previously. Each independent variable is split into $m = 100$ categories, each comprising $k = 100$ units. Random noise is added to the real-data percentages, drawn uniformly from (-.02, -.01, .01, .02).

For the non-linearity simulation, we fit two models to the data. The first is a logistic regression that fits a linear function of x_1, which is a wrong model for the data. The second is the correct logistic regression that fits a quadratic function of x_1. Figure 1 displays the perturbed percentages (labeled with $\circ$), and the predicted probabilities (labeled with $*$) from the incorrect model, plotted against the values of the midpoints of the categories of x_1. The predicted probabilities differ greatly from the perturbed percentages, so that the user should modify the model. As shown in Figure 2, adding the quadratic term to the model greatly improves the fit, and the predicted probabilities closely resemble the perturbed percentages. This clearly illustrates the utility of these diagnostics over releasing only coefficients and standard errors as the logistic regression output.

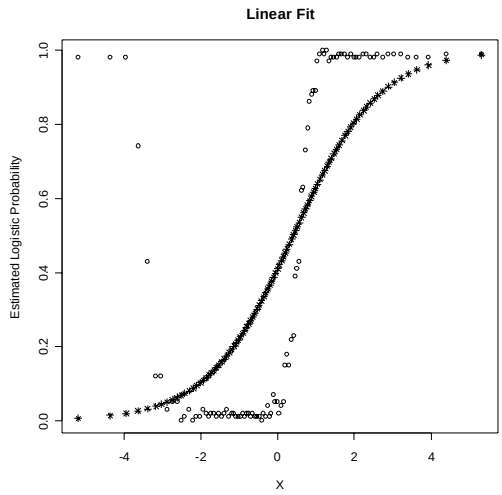


Figure 1: Synthetic diagnostics for non-linearity simulation when using the incorrect logistic regression model.

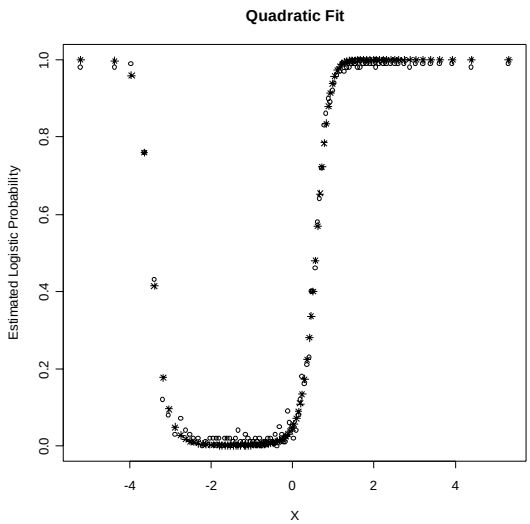


Figure 2: Synthetic diagnostics for non-linearity simulation when using the correct logistic regression model.

For the interactions simulation, we also fit an incorrect model and the correct model. The incorrect model uses just the main effects of x_1 and x_2, whereas the correct model includes the main effects and the interaction. As shown in Figure 3, the main effects model has poor correspondence between

the perturbed percentages and predicted probabilities; the diagnostics reveal that this model does not fit the data. Figure 4, on the other hand, shows coherence between the predicted probabilities and perturbed percentages, correctly suggesting that this model is a reasonable fit to the data.

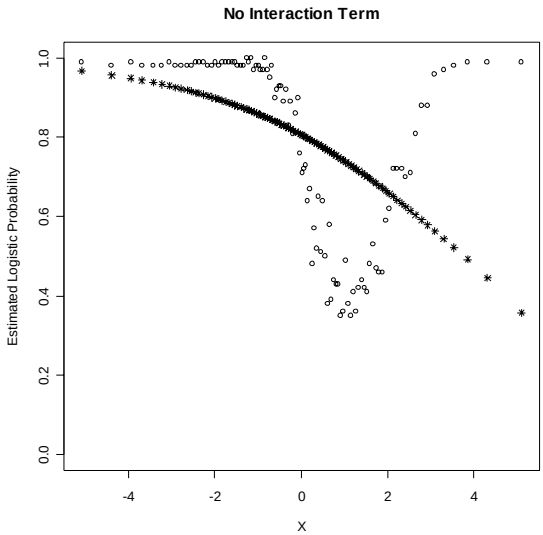


Figure 3: Synthetic diagnostics for interactions simulation when using the incorrect logistic regression model.

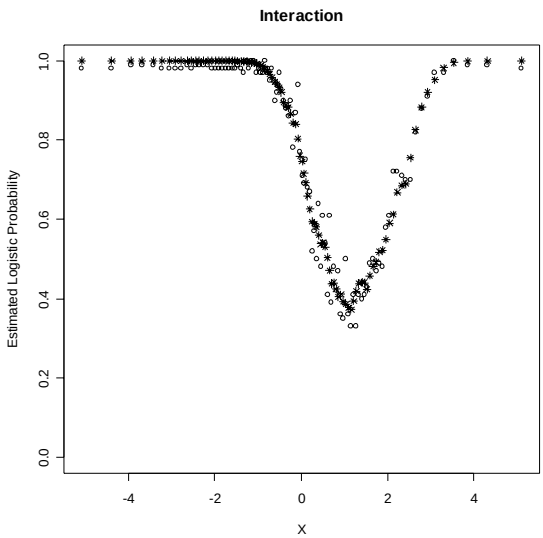


Figure 4: Synthetic diagnostics for interactions simulation when using the correct logistic regression model.

4. Remote Server Diagnostics for Multinomial Regressions

In multinomial regressions, the outcome variable has $h > 2$ categories. Each unit's residuals are an $h \times 1$ vector, $(y_1 - \hat{\pi}_1, y_2 - \hat{\pi}_2, \ldots, y_h - \hat{\pi}_h)$, where $y_j = 1$ when the unit's outcome is in category j and $y_j = 0$ otherwise. The only positive value in the residual vector is for the category matching the unit's outcome. Clearly, releasing even the sign of the unit's residuals from a particular multinomial regression query discloses the unit's outcome.

As we do for logistic regression, we propose to release grouped diagnostics rather than case-specific diagnostics. We partition the $\mathbf{x}_p$ into m_p categories, using at least 100 units per category when possible. For each of the m_p categories, we determine the number of units in each of the h outcome categories, and add random noise to these counts drawn from (-2, -1, 1, 2). The medians of the $\mathbf{x}_p$, the perturbed percentages of units in each of the outcome categories, and the average predicted probabilities are released so that users can determine whether or not the predicted probabilities closely resemble the perturbed percentages.

We illustrate these diagnostics using simulations like those in Section 3. We generate $n = 10{,}000$ observations for each simulation, with the independent variables, x_1 and x_2, randomly drawn from $N(0, 2^2)$. To generate the dependent variables, we draw from multinomial distributions with $h = 3$ categories and the following probabilities for the three categories:

$$\pi_0 = 1/(1 + \exp(g_1(x_1, x_2)) + \exp(g_2(x_1, x_2)))$$

$$\pi_1 = \exp(g_1(x_1, x_2))/(1 + \exp(g_1(x_1, x_2)) + \exp(g_2(x_1, x_2)))$$

$$\pi_2 = \exp(g_2(x_1, x_2))/(1 + \exp(g_1(x_1, x_2)) + \exp(g_2(x_1, x_2))),$$

with the values of $g_1(x_1, x_2)$ and $g_2(x_1, x_2)$ determined as in Table 2.

Table 2: Scenarios for multinomial regression simulations

Case	Functions of predictors
Non-linearity	$g_1(x_1, x_2) = 3.1 - 2.5x_1 + 1.8x_1^2$ $g_2(x_1, x_2) = 2 - 1.6x_1 + 1.4x_1^2$
Interaction	$g_1(x_1, x_2) = 2 + 1.5x_1 - 0.8x_2 + x_1x_2$ $g_2(x_1, x_2) = 1.5 + 2.2x_1 - 0.4x_2 + x_1x_2$

Using these data sets, we generate grouped diagnostics by splitting each independent variable into $m = 100$ categories, each consisting of $k = 100$ units. The real-data percentages for each category are perturbed by randomly adding random draws from (-.02, -.01, .01, .02). All multinomial regressions are fit using the "multinom" function in the software package S-Plus.

For the non-linearity simulation, we fit two functions of x_1 in the multinomial regression models: an incorrect linear function and the correct quadratic function. Figure 5 displays the perturbed percentages (labeled with $\circ$) and the predicted probabilities (labeled with $*$) from the incorrect, linear model. The predicted probabilities differ widely from the perturbed percentages, reflecting the lack of fit. As displayed in Figure 6, adding the quadratic term to the model greatly improves the fit.

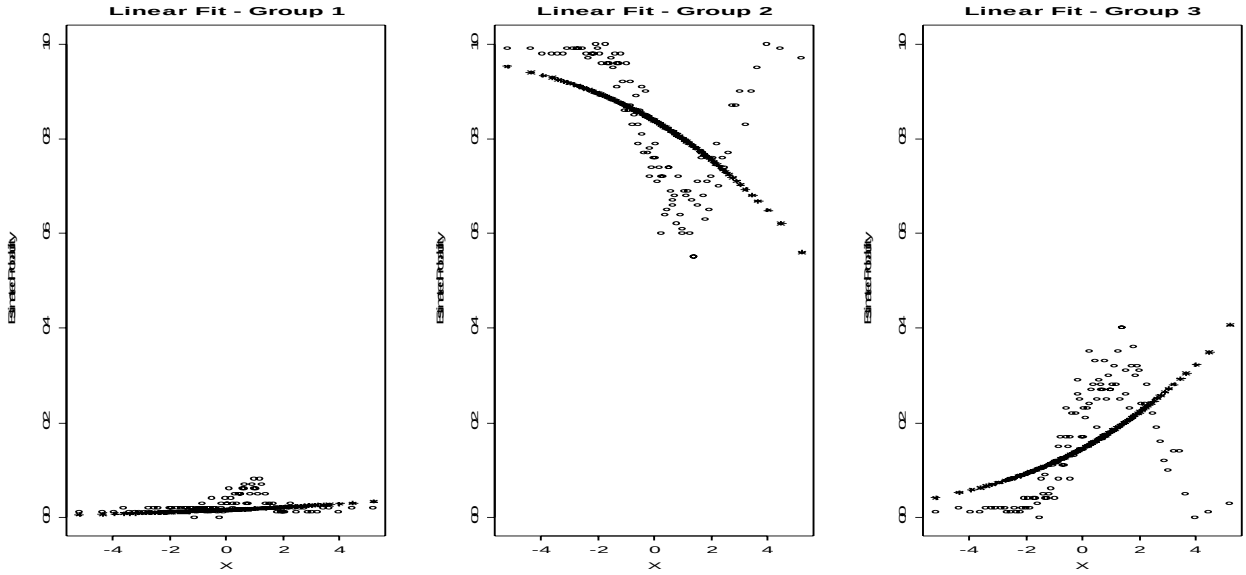


Figure 5: Synthetic diagnostics for non-linearity simulation when using the incorrect multinomial regression model.

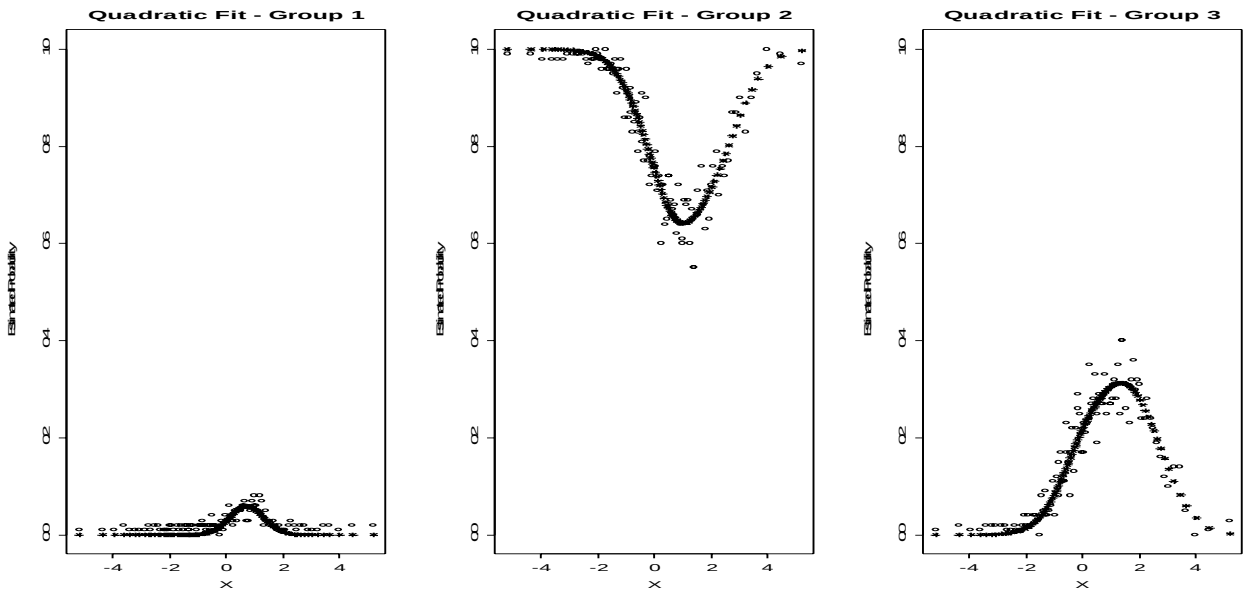


Figure 6: Synthetic diagnostics for non-linearity simulation when using the correct multinomial regression model.

In the interaction simulation, we once again fit both an incorrect and correct model. The incorrect model only uses the main effects of x_1 and x_2, whereas the correct model includes these main effects along with the interaction. The synthetic diagnostics in Figure 7 properly show the lack of fit for the main effects model, particularly for larger values of x_1. Figure 8 shows that once the interaction is included in the model, there is better correspondence between the predicted probabilities and perturbed percentages, which suggests the model is a better fit.

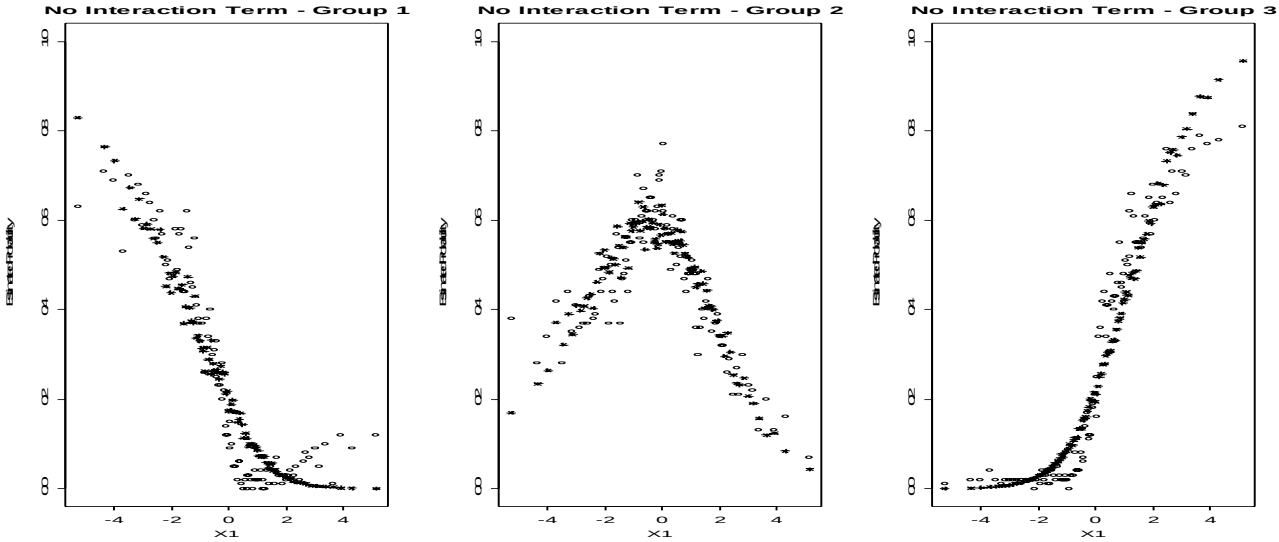


Figure 7: Synthetic diagnostics for interactions simulation when using the incorrect multinomial regression model.

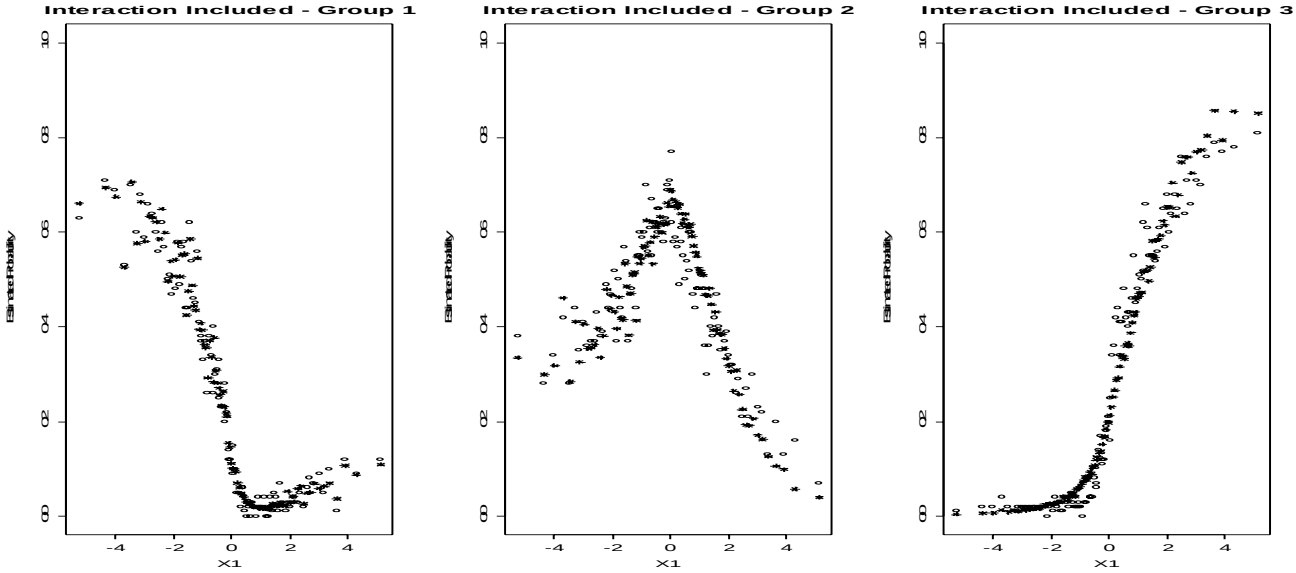


Figure 8: Synthetic diagnostics for interactions simulation when using the correct multinomial regression model.

5. Concluding Remarks

The synthetic diagnostics proposed here can improve the utility of remote access servers, without substantially compromising confidentiality, relative to releasing only the basic output from queries. There may be other ways for users to achieve disclosures from remote servers that are not addressed by these diagnostics. For example, when all units with the same covariate pattern have the same outcome values, the output from the fitted logistic or multinomial regressions allow users to predict those outcomes exactly. Furthermore, there may be judicious transformations of variables that allow users to determine certain outcomes exactly, for example fitting indicator variables that equal one only

for certain units. Decision rules for what to release, and how these rules interact with the release of synthetic diagnostics, is an area for future research.

References

Duncan G T and Mukherjee S (2000) Optimal disclosure limitation strategy in statistical databases: Deterring tracker attacks through additive noise. *Journal of the American Statistical Association,* vol. 95, 720-729.

Fuller W A (1993) Masking procedures for microdata disclosure limitation. *Journal of Official Statistics,* vol. 9**,** 383-406.

Hastie T J and Tibshirani R J (1990) *Generalized Additive Models.* Boca Raton, Florida: CRC Press.

Hosmer D. W. and Lemeshow S. (2000) *Applied Logistic Regression.* New York: John Wiley and Sons.

Keller-McNulty S and Unger E A (1998) A database system prototype for remote access to information based on confidential data. *Journal of Official Statistics,* vol. 14**,** 347-360.

Pregibon D (1981) Logistic regression diagnostics. *The Annals of Statistics,* vol. 9**,** 705-724.

Reiter J P (2003) Model diagnostics for remote access servers. *Statistics and Computing,* forthcoming.

Schouten B and Cigrang M (2003) Remote access systems for statistical analysis of microdata. *Statistics and Computing,* forthcoming.

Willenborg L and de Waal T (2001) *Elements of Statistical Disclosure Control.* New York: Springer-Verlag.

About the Authors

Jerome Reiter is Assistant Professor of the Practice of Statistics and Decision Sciences, and **Christine Kohnen** is a graduate student, at the Institute of Statistics and Decision Sciences of Duke University. Please address all correspondence to Jerome Reiter at Box 90251, Duke University, Durham, NC 27708, USA; tel. 919-668-5227; fax 919-684-8594; e-mail jerry@stat.duke.edu

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Apples and Pears. Techniques for Integrating Information

Steve Harris

Abstract

Like other organisations Market Research companies inevitably contain disparate systems using disparate hardware and software. In Market Research there is often a stark contrast between the systems used for storing information such as questionnaires, written reports, presentations, project plans, fieldwork reports, tables, graphs and respondent level data. This can often include many of the following storage formats: databases, word documents, PowerPoint documents, spreadsheets, plain text documents and specialised data formats. There is an increasing need for companies to make this valuable information more accessible, both internally amongst research departments and externally to clients.

Within agencies, service departments need to communicate information about the status of a project to research departments who in turn need to communicate research findings to the client. Because there is often no common ground between the software systems begin used, in the past it has been challenging to bring this information together.

This paper examines solutions for tackling these issues and will consider the use of open standards such as XML and Web Services, recent technologies that provide a useful mechanism for 'black boxing' software systems, allowing information from disparate systems to more easily be integrated.

The paper will explore a service-based approach as part of building a web based reporting system for BMRB. This system pulls together existing reporting software applications and existing project management systems into a client data portal. The incorporation of industry open standards such as TabsML from the OpenSurvey group is also considered and demonstrated, as well as interfacing with the SPSSmr Ltd. Dimensions library.

Keywords

Analysis Software, Reporting, Data Portal, XML, Web Services

1. Background

Single source information

A current buzzword in many industries is Portal. A Data Portal or Client Portal represents bringing together information from a variety of sources into a single place for fast, easy access, creating a single point of entry. In market research terms, for the client company, this typically means a single source of information about a given research project.

There are clear benefits for clients in having access to their research findings via a Portal:

- **Ease of use** – all of the information is in a single place.
- **Speed of response** – research data and documents can be published as soon as they become available, in some cases this may even be in real time.
- **Greater accuracy** – users can be more confident that they are looking at the most up to date information.
- **Archive** – previous research projects can be kept on-line for future reference.

There are also clear benefits for the research company:

- **Internal communication** – all of the information is in a single place, research and operations departments can be sure of using the latest most up to date information.
- **Ease of communication with the client** – publishing documentation and research data is made simple, can be automated and is consistent.
- **Improved service to clients** – a well designed client portal providing a one-stop-shop for all information relating to a research project, should give clients more reasons to use the research company again in the future.

It is clear that interest in Client Portals within research is high and they are now beginning to emerge in a number of research companies. Indeed a recent poll in Research magazine placed Portals second on the list of software development projects most likely to be worked on during 2003.

Whilst Portals are unlikely to revolutionise the industry, they will play a valuable role in streamlining processes and improving the quality of service.

The Challenge

Creating a Portal has in the past represented a fairly major technical challenge. Information is often stored in diverse data formats and disparate systems, which are incompatible. Often these systems may reside on different platforms and hardware, presenting further boundaries to integration. There is often no common ground, no common formatting, no common transport mechanism and unstructured data, in short no common way to send data back and forth.

This paper focuses on recent technologies, which help to address these problems of integration, without requiring systems to be tightly coupled. Tightly coupled systems with high levels of dependencies make any future change to the system or the data source hard to accommodate, reducing the usable life of the system. We'll come back to this subject later in the paper.

The Internet

The growth of Internet technologies provides a common user interface model, making the task of bringing information together more feasible. Company Intranets and Extranets were the first systems to harness this technology for this purpose. However, whilst HTML (Hypertext Markup Language – the format used for creating web pages) and the 'universal' browser is a powerful tool for presenting all kinds of information, it does not provide common protocols for exchanging information between systems. Allowing systems to 'talk' to each other was still a tough task. Intranets are also often very tightly coupled to the data sources that feed them, in some cases to the point where the information is authored directly in HTML.

2. What is needed?

Data Format Standards

Information comes in all shapes and sizes, from relational database systems, to MS Word and MS Excel files, to plain ASCII text files. None of these formats are ideal. Some are proprietary, some are hard to interpret, and some require specialist knowledge and software tools. We really need a data format that everyone can agree on. Such a format would ideally have the following qualities:

- Applicable to all kinds of information.
- Highly structured, so it is easy to interpret and manipulate.
- Human readable (verbose) and easily understood.
- Simple and not requiring specialist software tools.
- Open so it can be used by everyone.
- Adopted as an industry wide standard.

It is with these aims in mind that XML (Extensible Markup Language) was devised. XML is based on SGML (Standard Generalised Markup Language), the basis for HTML and a precursor to XML. Like HTML, XML is a simple text based format that uses tags to represent data; a simple example is shown below:

```
<?xml version="1.0" encoding="utf-8" ?>
<Projects>
   <Project id="gb02wi">GB TGI Winter 2002</Project>
   <Project id="gb03sp">GB TGI Spring 2003</Project>
</Projects>
```

This example demonstrates most aspects of XML. Even without knowledge of the XML format it is quite clear that this data represents a list of projects and associated ids. This is the power of XML, it is simple, easy to read and the tags mean the data is self-describing. The tag names themselves are not fixed and are defined for the specific purpose, (hence the word extensible in XML). Schemas can be used to define the tag names, creating data format standards. This approach can be extended to all kinds of data; there are even XML schemas for describing pictures, such as Vector Markup Language (VML).

XML is governed by the World Wide Web Consortium (W3C), an independent body lead by Tim Berners-Lee (the founder of the world wide web) whose aim is to promote open standards across the internet industry. XML has been widely adopted throughout the IT industry as the format of choice for storing and communicating information. Even MS Office documents such as Word and Excel are moving towards an XML based format.

Agreed Methods of Communication

Allowing different applications to talk to each other is not easy, especially if the applications are running on different machines and in particular if they are running on different platforms, such as Windows and Unix. Existing technologies are platform specific and do not easily integrate, it is often difficult to provide a bridge from one to the other. Up until now, the IT industry has not agreed on how such a request/response mechanism should work. What is needed is an industry standard mechanism for linking disparate systems with an agreed (open) common specification. Ideally such a mechanism should be simple to implement and should use existing data format standards. What has emerged is a specification called SOAP (Simple Object Access Protocol), which is itself based on XML. SOAP

essentially defines the structure of a request/response message that can be sent to any other application running on any other computer across a network. Because SOAP is based on XML it is very suitable for accessing computers remotely across the Internet and is easy to implement on any platform.

Requests using SOAP can be made to computers running under almost any operating system or hardware running locally or remotely across an internal network or across the Internet.

Interface Layer

Having good standards is a significant step in the right direction, but another very important requirement is the ability to define an interface into an application. More appropriately we are really talking about information sources, so perhaps a better concept is to think of services.

A service-based approach allows us to think of our data sources as a service we can call upon to provide some information on demand. Services will tend to be simplified higher-level interfaces into our data. We can define these interfaces using the Web Services technology described in the next section. Diagram 1a illustrates how Web Services might fit in to a Portal framework.

This interface layer provides a level of indirection into our data, effectively decoupling the system using the information from the system supplying the information, protecting us from future change. If the implementation details of a given data source change - say the project management system is replaced - so long as the information required for the service still exists, the service can be upgraded to use the new system, without any impact on the users of the information (see diagram 1b). This is likely to be the case since the nature of the information is defined by the business needs and will not change simply because the mechanics have changed. In this respect Web Services can be used to provide the business logic layer of our architecture.

The Web Service interface can be thought of as a contract between users of the information and the suppliers. This becomes really powerful if the users and the suppliers are geographically separated and particularly in business-to-business transactions.

Diagram 1a **Web Services Layer on data sources**	**Diagram 1b** **Web Service layer protects portal from changes made to underlying data sources**
	

3. XML Web Services

The introduction of XML and Web Services using SOAP (Simple Object Access Protocol) provides a powerful, universal method for the exchange of information. XML is a highly structured simple text based file format, which is easily created and read by software as well as being easy to read by humans. Web Services are built around SOAP, which is a universal protocol for making requests to (often) geographically remote systems. Web Services can provide an external interface into information sources.

When it comes to bringing information together, Web Services can be a real asset. Our approach is simple: expose each information source as a well-defined Web Service and use XML to structure all data. We can then build an application such as a client portal calling on these Web Services as required. As the business changes we can accommodate new information sources by defining new Web Services and change existing ones as necessary. Because of the level of indirection provided by the Web Service any change is not propagated across the overall system, but is isolated in a single place.

In building such a system we can make appropriate use of standard data formats where they exist. For example the support of TabsML as defined by the Open Surveys group should make incorporating tabular reports from a new source into a client portal trivial. The development and adoption of these open standards is key to leveraging the full potential from this approach.

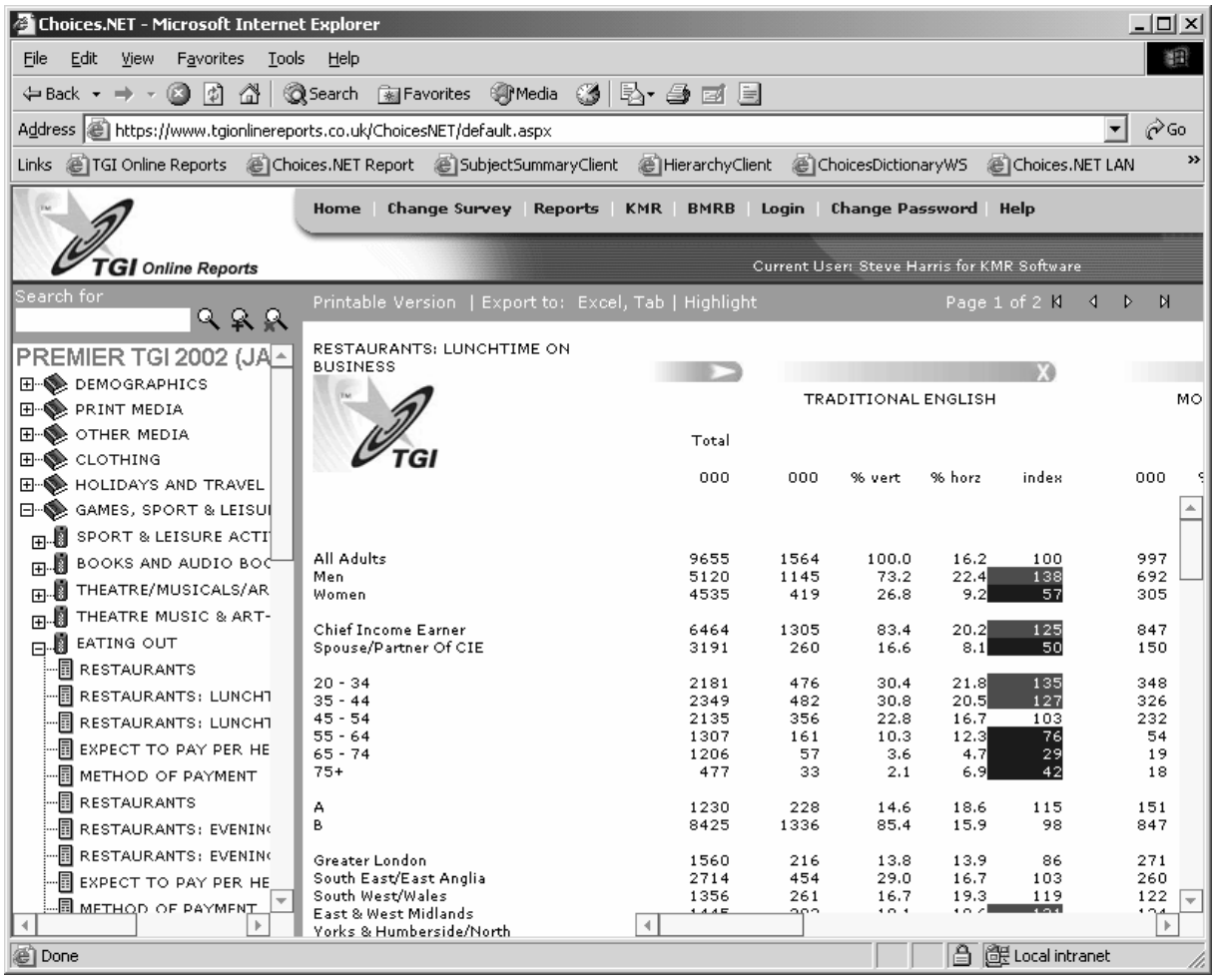
The next sections describe the development of two systems I have been involved in, which have followed this approach.

4. Putting it all together

Case study #1: TGI Online Reports

The GB TGI survey conducted by BMRB International has been running annually since 1969. In these early years the only published data was produced in a series of bound report volumes. Times have changed and the data are now primarily published via desktop analysis tools such as Choices 3 from KMR Software. However, the legacy report volumes are still produced. In Easter 2003 the report volumes were brought up-to-date and published online via the TGI Online Reports data portal, see diagram 2.

Diagram 2 – TGI Online Reports



Publishing the reports online offers many benefits. Reports do not need to be distributed either in hard copy or on CD-ROM. All clients can be assured of using the same set of reports. If a report needs to be reissued, it only needs to be changed once and all clients are updated. Over time a full central archive of reports will build up that can be instantly made available to new staff at client companies.

In developing this system, we took the decision to store all reports in a structured format using XML. The reports are pre-calculated for performance reasons, but there are some major advantages to using XML as the storage format. Traditionally the reports have either been printed or stored as plain text or RTF (Rich Text Format) files. These formats contain the entire layout formatting of the report making the information contained in the report difficult to access. By using XML we were able to separate the information from the formatting. Essentially the XML files contain all the information for a report and formatting is applied 'on-the-fly' using style sheets when the report is requested. This approach is very flexible and allows the reports to be published in multiple formats: from the XML we have been able to convert the reports to PDF Adobe Acrobat files, CSV files for Excel and Tab delimited files for other packages. This approach allows us to build up a library of report styles and formats over time. If in the future we want to change the format of these reports we can simply change the style sheet and the new format will apply to all reports available in the archive. XML has proved to be extremely versatile in this respect.

There are some issues with storing reports as XML though, the main one being size. XML is a very verbose file format and in many cases the entity tags can take up more space than the data itself. The

TGI reports are quite large tables, typically a couple of hundred rows and 5-6 columns. The XML files for these reports can be a few hundred Kbytes, this can present some challenges when converting to other formats, however, the XML parsers are improving all the time and using XSLT (Extensible Stylesheet Language Transformations) we have been able to overcome these issues.

The XML schema used is similar to TabsML from the OpenSurvey group. We had created an internal schema, which turned out to be very similar to TabsML in overall structure; however, we then made some modifications to bring it closer to TabsML. The Reports XML schema differs from TabsML in that it supports multiple filters for a table and multiple surveys for trending across survey waves. There are also some specific tags relating to information particular to our Choices software, such as the expression used for the column, row etc. Our reasoning for using our own schema is that we were looking to create a schema we could use throughout our Choices.NET software and we needed support for these extra features. However, by closely aligning this with the TabsML format we are able to import/export from TabsML very easily. Having said all of that we are also able to support TabsML reports on the site, which it is hoped will open up the functionality beyond the TGI surveys. A simple example of Reports XML is given in appendix I.

Having created our XML schema for the storage of the reports the system is built around a number of Web Services. In particular there are 2 key Web Services:

Reports web service – essentially this is a library of reports and associated documentation. This is the heart of the system and provides an XML-based dictionary of available reports for a given survey and the XML reports on demand.

Subscriber web service – this is a vital part of our online offering and controls access to surveys and reports via a client subscription database. Again this is provided as an XML based Web Service, providing lists of surveys clients have access to and validating users.

The use of XML Web Services to provide these reports online has simplified the task greatly and we have been able to extend both of these services for building a client portal (see case study #2).

Case study #2: Building a client portal

BMRB International has a number of excellent project management systems, field systems, client contact management systems etc., however, they are not all written in the same software and do not have web based user interfaces. Some are closed systems and some are expensive to alter. It is not possible for example to view field information from within the project management system. This could be added but at a high cost. Doing so would tightly couple these two systems and if the field system changed in the future so would the project management system have to change, and what about viewing the project management information from the field system?

The idea of a client portal is that all the various sources of information relating to a given research project should be accessible in a single integrated place. So which of the available systems do you base the portal around? Since a key aim of such a project is remote access and since none of the systems were web based, a new web system was built specifically for the portal. This is based around the same system used for TGI Online Reports: Choices.NET. XML Web Services were used to define interfaces into each of the data feeds. This protects the portal from future change to any of the underlying systems. If BMRB chooses to change its project management system, only the Web Service feeding the portal needs to change not the portal itself.

The Portal does not need any knowledge about the underlying structure of the information, only the XML schema and the location of the Web Service. These 2 when combined provide a documented interface into the information source, which can remain fairly static over time.

Many of the inputs and outputs for a research project are stored in general documents, (Word, PowerPoint, etc.), however, these documents are not all stored in the same location on the network. Once again we were able to create a Web Service to provide a virtual project folder to feed this information to the portal. The underlying network structure will no doubt change over time but the virtual project folder can be maintained.

The final information source for the portal is the raw respondent level data. Whilst databases such as SPSS SAV files can be made available for download just as any other document, we have used the Choices.NET application to provide an online analysis system. Choices.NET is built around the Choices 3 software analysis engine, which has its own proprietary data formats. However, we have adapted Choices to use the Dimensions library from SPSSmr, which simplifies the task of linking in data from ad-hoc surveys. The Dimensions Data Model provides another level of indirection in the overall system; data from a variety of different formats can now be made accessible via Choices.NET and the portal.

5. Conclusions

The advent of XML Web Services is ideally suited to the requirements of market research companies looking to integrate information sources across the business. This paper discusses the merits of this approach and demonstrates the direct application of this technology in providing 2 major web systems for clients. The incorporation of industry standards further enhances the flexibility of systems built around XML. XML Web Services can easily be adapted as the business needs evolve. The use of the services can also be extended to other applications and new services can be added as required.

Appendix I – Reports XML Example


```
<?xml version="1.0"?>
<Reports>
   <Report id="r1">
      <Name>GIN PRODUCT TABLES</Name>
      <DataTypes>
         <DataType type="1">
            <Label>000</Label>
         </DataType>
         <DataType type="2">
            <Label>% vert</Label>
         </DataType>
      </DataTypes>
      <Headers>
         <Title>
            <LeftText>Choices 3 Report</LeftText>
            <LineIndex>1</LineIndex>
         </Title>
      </Headers>
      <Footers>
         <Title>
            <LeftText>Sample Footer</LeftText>
            <LineIndex>1</LineIndex>
         </Title>
```

```
      </Footers>
      <Filters>
         <Variable id="1">
            <Value id="1">
               <Name>Adults 18+</Name>
               <Expr>DBAGEA18</Expr>
            </Value>
         </Variable>
      </Filters>
      <Columns>
         <Variable id="1">
            <Name>Gin:</Name>
            <Value id="1">
               <Name>Heavy Users</Name>
               <Expr>( BSGIF1OM OR BSGIF5T9 OR BSGIF3O4 OR BSGIF12W )</Expr>
            </Value>
          …
         </Variable>
      </Columns>
      <Rows>
         <Variable id="1">
            <Name>Age</Name>
            <Value id="1">
               <Name>18-24</Name>
               <Expr>DBAGE182</Expr>
            </Value>
          …
         </Variable>
      </Rows>
      <Survey id="gb02sp">
         <Source>GB TGI 2002 Spring (January 2001-December 2001)</Source>
         <Copyright>Copyright BMRB International 2002</Copyright>
         <Cell type="1" filterid="1" colid="1" rowid="1">
            <DataValue>3948.14</DataValue>
         </Cell>
         …
      </Survey>
   </Report>
</Reports>
```


References

Tabor, Robert, 2002. *Microsoft .NET XML Web Services*. Sams.

Berners-Lee, Tim, 2000. *Weaving The Web*. TEXERE Publishing.

Sunblad, Sten and Sundblad, Per, 2000. *Designing for Scalability with Microsoft Windows DNA*. Microsoft press.

http://www.w3c.org Main World Wide Web Consortium (W3C)

http://msdn.microsoft.com/xml

http://www.opensurvey.org OpenSurvey organisation site

About the Author

Steve Harris is Technical Director of KMR Software, a division of BMRB International that specializes in supplying software for the Market Research and Media industries. Steve joined the TGI division of BMRB International in 1987 having graduated from Reading University with a 2.1 in Zoology and soon after completed an MSc in Computer Science from Birkbeck College, University of

London. Steve has been responsible for numerous software and IT related projects at BMRB and has principally been responsible for the development of the Choices survey analysis software.

ASC 2003. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

The Impact of a Data Warehouse on the Survey Process

Mickey Yost

Abstract

Easy access to large collections of historical survey and census data, and the associated metadata that describes it, has long been the goal of researchers and analysts. Many questions have gone unanswered because, the datasets were not readily available, access was limited, and information about the business metadata was inconsistent, not well defined, or simply unavailable. This paper focuses on the impact of a Data Warehouse on the survey process including statistical sampling, survey methodology, data analysis, data quality, and data management. Survey processes at the National Agricultural Statistics Service (NASS) have traditionally focused on program areas that were largely self-contained. This meant the data collected for that program area was also self-contained. Data standardization, cross program analysis, strategic data management initiatives, and integration were considered quite impossible. In fact, there was no clear mandate (political capital) to push for such things - the individual program areas could see no advantage that would motivate changes in their program areas. This paper explores the strategic opportunities that came from the work of combining data from many different program areas into a single and integrated database. It will discuss the conditions that existed prior to the construction and implementation of a Data Warehouse and how the NASS Data Warehouse fulfilled one of its most important goals, the creation of the political capital needed to muster the reengineering and integration of our survey and census processes. It will discuss the changes that occurred in new systems development, data quality, analysis, collaboration across program areas, and survey execution. The paper will draw on previous papers that describe how to build a Data Warehouse, and what some of the design considerations should be, but will not deal extensively with these topics.

Keywords

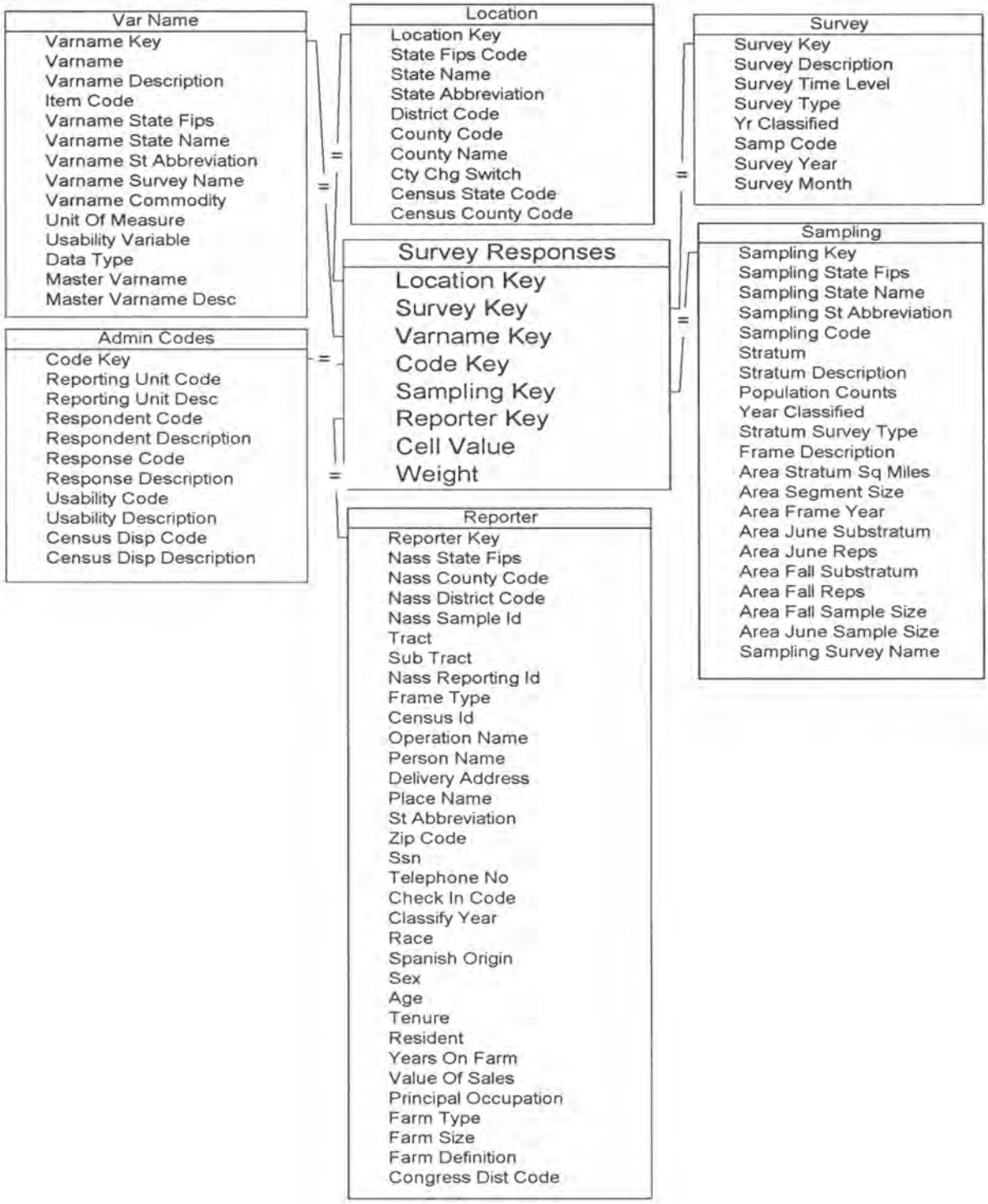
Data Management; Data Quality; Process Reengineering; Metadata; Integrated Data Sources; Database Design

1. The Data Warehouse and The Census of Agriculture

In 1996, the National Agricultural Statistics Service (NASS) began work on an easy-to-understand and easy-to-use high performance historical Data Warehouse (Yost, 1999, p. 376). At a minimum, the Data Warehouse would track historical survey and census responses from data providers, and easily manage changes in survey and census metadata over time. The data contained in it would also be readily accessible by all NASS employees for analysis and planning. It would also answer the

Agency's strategic need to continually improve statistical analysis, and survey and census procedures. The NASS Data Warehouse was

Figure 1



modelled using a dimensional design. This approach, unlike traditional modelling approaches used for transactional processing, studies the source data that will populate the Data Warehouse and identifies the information present on the data sets that can serve as dimensional views to the data. The choice of the dimensions is almost always done with an eye on the business reasons for analysing the data

(Adamson & Venerable, 1998, p. 464). For a statistical agency that conducts surveys via questionnaires, the dimensions were readily identifiable: Survey, Reporter, Location, Var_Name, Sampling, and Admin_Codes. See **Figure 1**. It is interesting to note the choice of dimensions was done by the analysts/statisticians, and not the technology professionals. Indeed, the NASS Data Warehouse was designed and built by the business units with little involvement from our Information Technology Division (ITD).

Our Data Warehouse went on-line in April 1998 with reported data from our core survey programs covering Crops and Livestock, and data from the Census of Agriculture, newly acquired from the Bureau of the Census. There was no attempt to back fill for years prior to 1997. The impact of the dimensional model was felt almost immediately. Analysts could access their data in ways not possible before, through a set of dimensions built to model their business rules. The analysts were given access to the actual database tables. The database tables were designed to be approachable, understandable, and easily remembered (Kimball, 1996, p. xxi).

I mentioned the Census of Agriculture was newly acquired. When analysts began to compare their survey and census data, the first important discovery was the sheer mass of data coming from the census reports compared to our surveys. Some, prior to this discovery, imagined the survey data would be used to true up any inconsistencies found in the census reports. That notion was immediately dropped. In a sample survey, much attention is paid to each individual report because of the potential for outliers and other sampling and non-sampling issues associated with a sample survey. This is done with a keen eye on the power of the sample design and the associated expansion factors. Individual reports stand for many other reports not present in the sample and thus have a much greater influence, record for record, in the estimation process. In a census, that leverage is not present and every report must be prepared to add, **not expand**, to the higher levels. The Data Warehouse, by demonstrating this key difference to the rank-and-file analyst, saved those analysts many hours of manual editing and data confrontation work that in the long run would likely have made little difference in the final tabulations.

I should point out here that all data currently present in the NASS Data Warehouse are for internal use only. After the Census was completed, there were, and still are, numerous special tabulation requests, made through official channels, that the Data Warehouse has been able to fulfil.

2. The Data Warehouse and Survey Data

After the Census data had been loaded, work continued on loading survey data. Applications began to be developed that would take advantage of the easy access to our census and survey data. The focus was on contextual access. For example, a Data Warehouse button was added to our List Sampling Frame Maintenance program that would pass a reporter's ID, the one being looked up, to a query against the Data Warehouse. The query would bring back to the desktop the known reporting history for that ID. The results took only seconds to display even though the number of reports in the Data Warehouse were accumulating at an astonishing rate. Currently there are over 1.8 billion rows in our central fact table taken from over 400 surveys, 7740 distinct survey questions, and nearly 2 million data providers whose reporting histories are on-line back to 1997.

Plans are underway to add the same functionality to our Computer Assisted Telephone Interviewing system. Issues had to be worked out on the proper use of historical data in the interview process to prevent inappropriate use of the data, such as suggesting an answer to speed the interview along.

Besides detailed information at the provider level, the Data Warehouse is capable of producing profiles, or higher level aggregates, that help the enumerator size up a respondent without giving the lowest level of detail. New applications to help get data into the Data Warehouse were also being developed. Our State offices needed to have fast and easy access to data collected and administered at the local field office, and applications were developed to facilitate loading their data into the Data Warehouse. As the impact of the Data Warehouse began to take effect, the strategic planners began to call for greater and greater use of the data for sampling, edits, imputation for missing data, analysis, and estimation. The surge in analytical power brought on by the Data Warehouse has created one of the largest reengineering efforts ever undertaken at NASS.

But before I can fully explore this effort, I must digress back to some of the lessons learned from the Data Warehouse.

3. The Data Warehouse and Metadata Quality

Early data sources used to populate the Data Warehouse were well behaved and consistent because they were coming from systems that represented our flagship processing systems. The data had been cleansed and stored with consistent assumptions about the data and what it meant. Metadata was present, consistent, and well documented for the data sources considered. The programs covered were also in the first rank of the Agency's product line, such as data covering Corn, Cotton, Soybeans, Wheat, Cattle and Hogs. One could not help but be successful with these data sources, the low hanging fruit, because they were consistently administered by a single system and by a single unit in our Headquarters.

Other program areas and processing systems were using some of the questions appearing in these core surveys. For example, "How many Acres of Land are on this place?" is a question on many of our surveys. Before the Data Warehouse, administration of these various surveys did not require metadata consistency, because there were no plans to combine data from different program areas together in one place. This is not to say the actual definition of the question was different across program areas, it was after all the same question. Rather, the metadata descriptions used in the various processing systems were different. These differences became a real issue as interest peaked and more and more program areas began to look for resources in the Data Warehouse to store their own data. The metadata for the All Land question mentioned above, for example, has as many names as there are programs that use the question. Clearly these inconsistencies had little impact across program areas until the data began to be loaded into the Data Warehouse. A report published by the Data Warehousing Institute in 2002, titled Data Quality And The Bottom Line, reported "data that today is deemed "sufficient to the task" or "good" may not be adequate to support future information needs." This statement was true in our case because each program using the All Land question considered their metadata adequate to the task. As the data were combined together in the Data Warehouse this was not the case. These inconsistencies were hampering our ability to do cross program analysis, and thus, to produce a higher quality product more efficiently.

The NASS Data Warehouse, by its dimensional design, created an important opportunity to get the metadata ***standardized*** (Yost, 2000, p.93) across program areas. This occurred because the variables are not the column definitions as in a wide flat file arrangement, but are stored as one row for each variable as in a long skinny design. Those occurrences of All Land each have their own row in the variable dimension table. By adding a column called Master Varname to the table (see example below), an alias was created that in effect tied all like variables together under one master name.

Legacy System Varname	Master Varname	Master Varname Desc	Master Survey Name
C001	CLANDTOT	LAND IN FARMS	COUNTY ESTIMATES SURVEY
C100	CLANDTOT	LAND IN FARMS	FRUIT AND NUT SURVEY
C900	CLANDTOT	LAND IN FARMS	TOBBACCO SURVEY
C901	CLANDTOT	LAND IN FARMS	OILSEED SURVEY
C995	CLANDTOT	LAND IN FARMS	VEGETABLE SURVEY
CLANDTOT	CLANDTOT	LAND IN FARMS	AG SURVEY
K47	CLANDTOT	LAND IN FARMS	CENSUS
FLAXCURR	CLANDTOT	LAND IN FARMS	HOG SURVEY
FNAXCURR	CLANDTOT	LAND IN FARMS	CATTLE ON FEED
IC101	CLANDTOT	LAND IN FARMS	DAIRY SURVEY

The effect of doing this is to allow legacy systems to continue using the old variable and new systems to use the new. As the legacy systems are updated or reengineered, the Master Variable is used until all systems are on the same standard. Moving to a consistent standard for metadata has some profound effects on data quality and efficiencies. As systems are designed to take advantage of consistent and repeatable metadata, systems integration becomes a more achievable goal. Developers are encouraged to not reinvent the wheel, but to use the standards that have been implemented in the Data Warehouse. The analyst benefits greatly as well because an enterprise view of all the data is available at greatly reduced expense. The Data Warehouse and the dimensional design proved data from many different data sources could be integrated into a single database and maintained over time. A well integrated collection of historical data that offers a consistent and repeatable view of the business metadata can and should be the basis for a well integrated process environment. This is the basis for real advances in data quality because integration benefits execution and execution benefits quality.

4. The Data Warehouse And Reengineering

When the notion of a Data Warehouse was being investigated, a recurring theme was that a Data Warehouse would likely form the basis for the reengineering of old stand alone processing systems. Some of the reasons cited at that time included freeing the legacy systems from the responsibility of storing history. Pre-Data Warehouse production systems offered little access to historical data mainly because of size and efficiency constraints. Also, allowing ad hoc access to production systems was not recommended. These systems managed to dodge the issue of integration as well because the need for standard metadata was not the primary concern of these systems. The goal was to develop the specifications and metadata, code the parameters, and publish the results. Strategic and cross program analysis suffered because combining data from many different data sources, each with their own metadata definitions, was problematic at best.

Clearly the old processing systems needed to be redesigned in order to take advantage of the rich history now available through the Data Warehouse. In 1998, NASS began work on a completely new system that would process both the 2002 Census of Agriculture and our regular on-going survey programs. By process, I mean a system that would at least handle survey specifications, editing, imputation, analysis, administrative reporting, summarizations, and publications. The new system would be designed to take full advantage of the data contained in the Data Warehouse by allowing direct comparisons between current and historical data for real time data editing, imputation, and analysis. The model developed to handle these requirements is known as an Operational Data Store (ODS). The ODS shares many of the same dimensions used in the Data Warehouse, such as

Var_Name, Reporter, and Location, which greatly enhances the comparison of current and historical data through these shared dimensions. Currently, this system is being used to process the 2002 Census of Agriculture.

To carry this one step further, the strategic planners are now proposing an architecture that will store survey and census specifications (a specifications repository) in such a way as to make the **sharing** of this important information easy and seamless across program areas. Meetings to discuss the nature of this architecture were held, and an important conclusion was reached. The actual elements of information contained in this new data architecture and their definitions would have to be agreed upon through a series of meetings over a period of several months. The questions that need to be answered have been some of the most elusive in the history of our Agency. What data elements or metadata are critical to many of our programs and thus should be shared, and what elements should remain exclusive to the program area and not be shared? The prospect of knowing what data elements are important to many programs and creating a single and shared repository for them opens the door to improved program execution and better data quality. It will also create opportunities for shared collaboration across all program areas using the repository. Also, I am sure there will be some data elements and processes currently in place, and using resources, that are really not necessary because it will be determined that nobody really needs the stuff.

5. The Data Warehouse and the Organization

I have been working up to perhaps one of the most important changes the Data Warehouse has brought to NASS. The dimensional design for integrating data is really a metaphor for integrating work processes and sharing information and knowledge. The nature of the design, a business-driven model rather than a data-driven model, taught us an important lesson: if we can integrate our data we can integrate our work processes and our programs. Before the Data Warehouse, programs devoted to collecting data for the Livestock Program, or the Crops Program, or the Economics Program were planned and executed independently of one another. This allowed for efficient execution of the program, but sealed off access to the key data values that could reasonably be shared across the program. The clear strategic advantage of having all of the collected data from a set of state and national farm surveys created a force for change in our organization and alignment. The first change came with the appointment of a program manager for the Census Program. This person's responsibility included bringing people from different programs throughout the Agency together as a team, in effect reporting to the new program manager. This was a departure from the traditional chain of command that was largely based on function rather than program. The efforts by the team were largely focused on execution of the combined functions. In analysing the relationships between each of the functional units, new critical paths were discovered that allowed for more effective load balancing between the functions, thus dramatically improving performance. Not to lose the thread of the original argument, but this is exactly why a dimensional model is so effective in delivering up strategic information. When queries are posed to the model the dimensional perspective **shares** information about how the dimensions interrelate cutting query times significantly.

6. Summary

Our Data Warehouse has had a profound impact on the way we do business. It has changed the way we analyse data, created opportunities to better manage our customer relationships, and forced us to

consider the impact of poor data quality on our product line. It has also taught us the importance of business and technology alignment. When the project started, support and development came mainly from the business units with little support from our IT division. The IT staff that would normally have worked on this were busy working on something called the Y2K problem. The business units, on the other hand, were in a great deal of pain from having insufficient access to a respondents history, or the capability to summarize this data in new and revealing ways. This fact helps explain the project's success because of the broad and high level support coming from the business units. It also, however, explains some of its failings. Without strong IT involvement, applications development has lagged behind expectations and integration with current systems has been slow. In spite of this the compelling impact of real data integration and the convenience of having the lion's share of the Agency's data on-line has created sufficient political capital to form new partnerships that did not exist before between IT and the business units. Douglas Hackney, in a keynote presentation at The Data Warehouse Conference 2002, sponsored by DAMA North East said, "Each Data Warehouse/Business Intelligence system that contributes value attracts a political constituency." As usage climbed and more functional units began to use the Data Warehouse, the impact of integration began to be measured and talked about. Developers in some of our IT units, wanting to create applications that would use the Data Warehouse, started to become a voice in favour of what was being accomplished. Indeed, the movement to re-architect to a single metadata standard could not have been done without these new alliances.

Data integration using the Data Warehouse has changed almost every critical path we use for collecting and analysing survey and census data. It has created new processes and capabilities for getting the job done as well as new and innovative technologies that leverage this intelligence. It has shown us that not only has technology changed, but our organizational structures and alignments have also been impacted. Currently there is a plan before top management to reorganize our IT division. A new branch is being proposed to consolidate all data management functions, and to refocus our applications development to take advantage of this new organizational structure.

The Data Warehouse has focused our attention on the relationships between our business processes and the data collected through those processes and is helping to drive the changes necessary to address those issues. These relationships in turn are focusing our attention to an enterprise view of our architecture - a view that includes our business functions, program requirements, applications, administrative data, and infrastructure. In the years to come, the Data Warehouse will be seen as the foundation for a new enterprise architecture that maximizes data integration, quality, and efficiency.

References

Adamson, C., Venerable, Michael (1998). Data Warehouse Design Solutions, New York: John Wiley & Sons, Inc.

Kimball, R. (1996) The Data Warehouse Toolkit. New York: John Wiley & Sons, Inc.

Nealon J. (2000) Improving Our Agricultural Statistics Program Through Easy and Fast Access By Employees to Previous Survey and Census Data. Staff Report, National Agricultural Statistics Service, 2000

Yost, M. (2000) Using a Dimensional Data Warehouse to Standardize Survey and Census Metadata. Proceedings of The1999 Federal Committee on Statistical Methodology Research Conference, 2000

Yost, M. (1999) Data Warehousing and Decision Support at the National Agricultural Statistics Service, USDA. Proceedings of The Third ASC International Conference, 1999

About the Author

Mickey Yost is a Mathematical Statistician with the National Agricultural Statistics Service in Washington, D.C., and currently serves as the Data Warehouse Architect and Manager for NASS's Data Warehouse System. He is recognised as a leader in database designs for data warehousing, and has conducted numerous training and consulting seminars on tracking survey and census data using data warehousing architectures. He can be reached at 1400 Independence Ave., SW, Room 6312-S, Washington, D. C. 20250-2010, myost@nass.usda.gov.