Publisher and Library/Learning Solutions (PALS)

Pathfinder Research on Web-based Repositories

FINAL REPORT

The institutional repository, an open web-based archive of scholarly material produced by the members of a defined institution, has come to the fore following the launch of DSpace at MIT at the end of 2002. This report reviews recent developments and quantifies the growth of institutional repositories, and explores the impact their expansion may have on scholarly publishing.

Mark Ware Consulting Ltd

Publishing and Elearning Consultancy

14 Hyland Grove
Westbury-on-Trym
Bristol BS9 3NR
UK

Telephone: +44 117 959 3726
Fax: +44 117 959 3726

e-mail: mark@markwareconsulting.com
Web: www.markwareconsulting.com

January 2004

Contents

1 Executive summary

An institutional repository (IR) is defined to be a web-based database (repository) of scholarly material which is institutionally defined (as opposed to a subject-based repository); cumulative and perpetual (a collection of record); open and interoperable (e.g. using OAI-compliant software); and thus collects, stores and disseminates (is part of the process of scholarly communication). In addition, most would include long-term preservation of digital materials as a key function of IRs. (Page 5)

The rationale for IRs is seen by some to be reform of scholarly communication, and in particular scholarly *publishing*; and also to enable the institution to enhance its prestige by making visible the fruits of its faculty's academic and research labours. More broadly, they are also seen as part of the digital infrastructure of the modern university, offering a set of services for the management and dissemination of digital materials created by the institution and its community members. (Page 6)

A convergence of technology developments and other initiatives has made IRs possible. Technology costs, especially storage costs, have dropped significantly. Standards such as OAI-MHP have been developed and developments in web publishing such as the open archives initiative have pointed to opportunities. There is now a variety of open source and commercial software platforms available for an institution wishing to develop an IR. (Page 7)

We have developed a list of some 45 IRs that formed the basis of our subsequent analysis. A number of the more prominent IRs and IR-related projects, including DSpace, EPrints, DARE, SHERPA, CODA, eScholarship and CARL, are described in Section 3. (Page 9)

The issues facing those establishing IRs are reviewed in Section 4. The most important are probably the cultural issues affecting faculty take-up of self-archiving services (e.g. inertia, lack of awareness, lack of a preprint culture) and intellectual property rights. Other issues include organisation and management, funding and business models, accession policies, metadata, long-term preservation and access. (Page 17)

In section 5 we discuss the main documented uses for IRs. These include scholarly communication; education; e-publishing; collection management; long-term preservation; institutional prestige; knowledge management and research assessment. (Page 20)

We review how established repositories are managed within institutions, and who is responsible for their maintenance and providing access to them, and also give some on the costs of establishing and maintaining an IR based on data from DSpace. (Page 22)

IRs need to develop policies for the deposit of materials. We look at allowed users; types of document (e.g. pre-prints, post-prints, technical reports, working papers, theses, etc.); supported digital formats; submissions approval processes; copyright policies; and policies on the removal of deposited papers (generally not permitted) and on compulsory deposit (e.g. for postgraduate theses). (Page 25)

To quantify the scale of IRs and the types of content, we look first at the wider set of about 250 OAI repositories (of which IRs are a subset), then at a list of 45 IRs developed for this research. Data on the 133+ repositories using EPrints software (which overlap with IRs) are then reviewed and finally we look at some data on academics' self-archiving behaviour at Edinburgh. (Page 27)

We find that IRs are currently rather small, with an average (median) of 290 records per institution (smaller but comparable to the median size of other OAI data providers). The quantifiable data reviewed supports the following broad conclusions (Page 33):

- the number of e-print repositories has grown reasonably quickly but there has been a problem in persuading faculty to populate them once launched. This is supported by accounts of those starting repositories;

- looking specifically at IRs, the majority are clearly in a very early stage of development but even most of the longer-established sites have a relatively small number of documents compared to the research output of their institutions;
- IRs to date have largely replicated the subject bias found in the older subject-based archives, i.e. content is largely maths, physics, computer science and economics;
- e-prints are currently a small fraction (~22%) of the content on IRs and post-prints currently appear to be a small fraction of these e-prints (though we have not established statistics on this split). However, this may be just a reflection of the early growth status of most IRs, or of a decision not to tackle controversial rights issues at first;
- there is therefore little support in this evidence for IRs leading the reform or disaggregation of scholarly publishing;
- it is not clear what IRs are currently adding to the long-term preservation agenda, given their patchy coverage.

A short survey of publishers was conducted to take a snapshot of their attitudes towards IRs. Some 45% think that IRs will have a significant impact on scholarly publishing, but almost as many (38%) don't know. Significantly, three-quarters think that the impact will either be neutral or there will be no impact. The stances towards IRs were split between "wait and see" (38%) and active experimentation (42%). Publishers appeared fairly relaxed about pre-prints but much more concerned about self-archiving of final published papers. Interestingly, while 55% permit authors to self-archive, some 12% said they currently permit but expect to tighten restrictions to exclude IRs. (Page 35)

We then review the issues raised for publishers by IRs. Copyright issues and the impact on journal subscriptions are obvious areas, but publishers are also concerned about the impact on journal submissions, future publishing models, the integration of IRs and journals and opportunities for collaboration. (Page 36)

In conclusion, the case for the benefits to a research organisation of an institutional repository providing a set of infrastructural digital services including uploading/hosting, organising (metadata), disseminating and long-term preservation seems compelling. Most universities of any substantial scale do now appear to be either implementing or considering implementing such a repository, and funding bodies throughout the world are supporting research into their development.

What is far less clear is whether IRs will develop large, interoperable collections of *published* literature, as hope the advocates of open access. IRs are currently at an embryonic stage with only small, experimental collections of documents, but a clear message from the IRs is that one major hurdle – possibly the major hurdle – is overcoming faculty's inertia or indifference to self-archiving. It seems possible at present that IRs *per se* will fulfil a real and valuable function in supporting scholarly communication, research and teaching but that this function will be complementary to scholarly publishing rather in conflict with it. The impact of the wider open access movement is of course another matter.

Publishers in our survey appeared to share this view, as the large majority believed that IRs would have either no impact or a neutral one on scholarly publishing. Nonetheless there are clearly substantial challenges both in dealing with the wider issue of open access, of which IRs form a part, and in responding to the specific opportunities and issues raised by IRs.

2 Developments

This report covers the development and present status of institutional repositories (IRs). IRs are not entirely distinct from other types of web repositories or archives, and there is considerable overlap with the Open Archives/Open Access movements, as we shall see later, but there is sufficient interest in IRs among the academic and library (and publishing) communities to regard them as a separate phenomenon. In the words of Clifford Lynch (Lynch 2003):

> *"The development of institutional repositories emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication, both moving beyond their historic relatively passive role of supporting established publishers in modernizing scholarly publishing through the licensing of digital content, and also scaling up beyond ad-hoc alliances, partnerships, and support arrangements with a few select faculty pioneers exploring more transformative new uses of the digital medium."*

2.1 Definition

Defining an IR is not an entirely neutral exercise, as individuals' views do vary depending on the role they see for IRs, particularly in relation to reforming scholarly publishing. Most if not all would however agree that an IR:

- is a web-based database (repository)
- of scholarly material;
- it is institutionally defined (as opposed to a subject-based repository),
- cumulative and perpetual (a collection of record),
- open and interoperable (e.g. using OAI-compliant software),
- and thus collects, stores and disseminates (is part of the process of scholarly communication).

In addition, most would include long-term preservation of digital materials as a key function of IRs. There are dissenters, however – mainly those who think that IRs should focus on freeing the research literature from subscription-model price constraints[1].

An IR can be contrasted with a disciplinary (subject-based) repository such as the arXiv or CogPrints repositories in physics and cognitive sciences respectively.

2.2 SPARC

The SPARC position paper on IRs (Crow 2002) has been influential in setting the agenda for the IR debate. Crow argues that there are two key rationales for IRs:

- reform of scholarly communication, and in particular scholarly *publishing;*
- to enable the institution to enhance its prestige by making visible the fruits of its faculty's academic and research labours.

For SPARC, repositories allow publishing to be disaggregated, separating the journal functions of registration, certification, awareness and archiving. In this regard, there appears little difference from other types of open web archives, except that Crow argues that the

[1] Note that "repository" has become the preferred term rather than "archive" because the latter implies the panoply of stewardship and long-term preservation that may not in fact exist.

institution has strong motives (reducing library budgets, institutional prestige) to create incentives for academic authors to archive their materials, thus overcoming the inertia that has so far prevented the large mass of authors from following physicists down the self-archiving route.

2.3 Lynch

Clifford Lynch, the Executive Director of the Coalition for Networked Information (CNI), published a thoughtful and generally unpolemical article in February 2003 (Lynch 2003). The subtitle, "Essential infrastructure for scholarship in the digital age", reflects Lynch's view of the strategic importance of IRs as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members." Lynch sees these services as fundamental and essential to scholarship in the digital age:

> *"At the most basic and fundamental level, an institutional repository is a recognition that the intellectual life and scholarship of our universities will increasingly be represented, documented, and shared in digital form, and that a primary responsibility of our universities is to exercise stewardship over these riches: both to make them available and to preserve them. An institutional repository is the means by which our universities will address this responsibility both to the members of their communities and to the public. It is a new channel for structuring the university's contribution to the broader world, and as such invites policy and cultural reassessment of this relationship."*

So for Lynch, the strategic importance of IRs lies in:

- remedying the weaknesses of current local self-archiving: running personal or departmental web servers is wasteful of academics' time and academics frequently lack essential skills (e.g. metadata creation and maintenance); it creates security holes (e.g. multiple poorly-secured websites connected to the institution's network); and it jeopardises the long-term preservation of materials because of lack of proper back-up procedures;
- providing a long-term solution: only academic institutions have the incentive and are in a position to address the issues involved in long-term preservation;
- improving scholarly communication: Lynch is keen to stress that he is interested in scholarly communication most broadly defined, not just scholarly publishing. He points to the development of new "born-digital" works that have do not (and cannot have) print equivalents, to the data-intensive nature of much scholarship, and asserts that

  > *"Journals will move too slowly and too unevenly to manage these resources, and disciplinary data repositories cannot be comprehensive. Institutional repositories can maintain data in addition to authored scholarly works. In this sense, the institutional repository is a complement and a supplement, rather than a substitute, for traditional scholarly publication venues."*

- Extending the work of disciplinary archives, which Lynch believes cannot be comprehensive;
- Improving teaching: not only by supporting campus teaching and "digitally captur[ing] and preserv[ing] many of the events of campus life—symposia, performances, lectures" but also by globally disseminating teaching via the web (e.g. MIT's OpenCourseWare initiative).

2.4 Why now?

There has been a convergence of technology developments and other initiatives that have made IRs possible. Technology costs, especially storage costs, have dropped significantly so that repositories are now affordable. Standards such as the open archives metadata harvesting protocol are in place and work is being done on the metadata itself. Awareness of the needs for, and challenges of, digital preservation has accelerated (e.g. the $100m grant to the Library of Congress). Developments in web publishing such as open archives initiatives, open access journals and disciplinary archives are pointing the way to opportunities to enhance scholarly publishing.

The launch of the DSpace IR at MIT in November 2002 (see MIT 2003 for an interesting case study of the launch of DSpace) and the subsequent release of the DSpace software under an open source licence have been seen as a seminal event and a trigger for the development of IRs. (DSpace is described in more detail below.) However the release of the EPrint software from Southampton University in January 2001 predated this by nearly two years and at present there are more functioning IRs running on EPrints than on the DSpace platform.

2.5 Software

There appears to be general agreement that there is now adequate, easily available software to create and maintain an IR. As a result, the challenges in setting up an IR are now seen as being less to do with technology (although the problems of long-term preservation are very far from solved) and more to do with managerial, organisational and cultural issues.

The two leading software packages, DSpace (MIT) and EPrints (Southampton) are both available free under open source licences, and there are at least half a dozen other possible packages. In theory, commercial document management or knowledge management software packages might also be suitable[2] but are unlikely to be adopted given their costs.

There is an (incomplete) Guide to Institutional Repository Software available at the Budapest Open Archive Initiative website (see BOAI 2003) which includes a feature-set comparison of DSpace, EPrints, DCSware, i-Tor and MyCoRe. (Note: URLs for organisations and websites mentioned in the text are given in Appendices 11.2 and 11.3.)

2.5.1 DSpace

The DSpace software has been purpose built in collaboration between Hewlett Packard and MIT to offer IR services. It is specifically designed to manage diverse heterogeneous types of digital content. It offers interoperability via OAI-MHP (Open Archive Initiative – Metadata Harvesting Protocol – a software standard that allows specialised search engines to gather article metadata from compliant websites) and built-in support for Dublin Core metadata (Dublin Core is an agreed metadata standard used in library cataloguing and elsewhere, though other metadata schemes are possible). It uses persistent identifiers (which are like URLs but have the benefit that unlike ordinary URLs they do not change when the linked document's physical location is changed) via the CNRI Handle system[3].

2.5.2 EPrints

The EPrints package is more oriented towards e-print archives, as the name suggests. It is also OAI-MHP compliant. It does not directly support persistent identifiers (though presumably it does not rule them out).

---

[2] For example, at one stage the Ohio State KnowledgeBank project was considering the commercial Documentum package (Branin 2002) although it has now decided to go with DSpace.

[3] See http://www.handle.net/ for further information

A comparison between the DSpace and EPrints packages has been published by William Nixon, a project manager at the DAEDALUS project at Glasgow University (Nixon 2003). He concludes that they have much in common and "is not a question of which software is better but rather which is appropriate for the institutional services which you are building, their purpose and the content. Will it be to free research papers or is it to manage and preserve digital content, or both?"

2.5.3 Other packages

Other software packages being used, or planned to be used, for IRs include (see Appendix 11.2 for URLs):

- **CDSware**: developed by CERN and used to run its very substantial CERN Document Server (over 630,000 bibliographic records, including 250,000 fulltext documents);
- **bepress**: created by The Berkeley Electronic Press for the University of California's eScholarship Repository;
- **Kepler**: The purpose of Kepler is to give any user the ability to easily self-archive publications by means of an "archivelet": a self-contained, self-installing software system that functions as an Open Archives Initiative data provider;
- **Fedora**: an ambitious project developed jointly by Virginia and Cornell with funding from Mellon. Fedora is a general-purpose digital object repository system that can be used in whole or part to support a variety of use cases including: institutional repositories, digital libraries, content management, digital asset management, scholarly publishing, and digital preservation;
- **i-Tor**: Tools and technologies for Open Repositories was developed by the Innovative Technology-Applied (IT-A) section of Netherlands Institute for Scientific Information Services. i-Tor acts as both an OAI service provider, able to harvest OAI compatible repositories and other databases, and an OAI data provider;
- **MPG eDoc**: developed by the Max Planck Gesellschaft in cooperation with the Fritz-Haber Institute. Currently used by many Max Planck institutes to "capture, document, share, archive, publish, disseminate and manage their scientific documents and the results of their research";
- **MyCoRe**: MyCoRe grew out of the MILESS Project of the University of Essen and is now being developed by a consortium of universities to provide a core bundle of software tools to support digital libraries and archiving solutions (or Content Repositories, thus "CoRe");
- **OPUS**: Online Publications University of Stuttgart. Also used by University of Konstanz;
- **Ebrary**: the aggregator / database company is offering a "new product that enables libraries to cost-effectively create online institutional repositories of documents such as theses and dissertations, technical reports, e-prints, articles, curricula guidelines and special collections. In preparation, we're extending a free pilot program to our existing customers and libraries that subscribe to our databases";
- **Ingenta**: have already announced a collaboration with Southampton University to develop a commercial version of EPrints. Ingenta say that they have undertaken considerable research into author/university requirements.

3 Institutional repository projects

A list of current IRs and their URLS is given in the table in Appendix 11.3. This list of some 45 IRs was produced using the SPARC website's list of IRs as a starting point, comparing it to the list of IRs on the EPrints.org website and adding additional sites that were discovered during this research.

To some extent, inclusion or exclusion of a particular site is a matter of taste but in compiling this list we have tried to include sites that best fit the definitions given above. For comparison, there are 243 OAI data providers covered by OAIster but most of these do not meet the full definition. OAIster-covered sites are analysed separately, however (see section 8.1).

3.1 DSpace

The DSpace project at MIT was funded by Hewlett Packard to the tune to $1.8m, plus 3 FTE HP staff and $400k in systems equipment. The DSpace project is well documented (e.g. see the DSpace website).

DSpace is the term used for the project (conceived in the 1990s, funded by HP from 2000); the longer-term research programme; the software developed by the project, which is now available under an open source licence; and finally DSpace is the name of MIT's IR.

The DSpace Federation is a group of institutions that are using DSpace software to build their own repositories. The DSpace Federation Project is a one-year study which will begin the process of building a collaborative federation of institutions running DSpace. This group will test the adaptability of the system to a targeted group of institutions with varied needs. Federation Project members include Cambridge, Columbia, Cornell, Massachusetts Institute of Technology, Ohio State, Rochester, Toronto, and Washington.

3.1.1 DSpace@MIT

DSpace@MIT was the first DSpace-based IR to be launched, in November 2002. The idea for a digital repository originated in the MIT libraries in 1997. The project received co-development funding from HP in March 2000. MacKenzie Smith was appointed the project director in January 2001. Early adopters (similar to beta testers) came on board in March 2002 with the full launch in November 2002.

DSpace is organised by Communities and Collections. A community could be any organisational unit and can support any number of collections. At the time of writing there were eight live communities as shown in *Table 1* on the following page.

The content is primarily grey literature (engineering technical reports and economics working papers). It seems likely that these collections have been added *en masse* from existing collections rather than submitted by individual authors (for example, the document dates go back to 1968). This is a very limited range of type of content compared to the remit of the institutional repository: where is the video, the datasets, the digital-only projects, the examples of "new digital scholarship" etc.?

A similar but independent project at MIT is the OpenCourseWare initiative, which aims to collect and make available MIT course materials.

Formats appear (from semi-random sampling) to be all or mainly PDF.

Dspace @ MIT

Communities	*Collections*	*No. records*	*Notes*
Center for Global Change Science	Joint Program on the Science and Policy of Global Change Reports	97	Series of Technical Reports
Center for Innovation in Product Development (CIPD)	Distributed Object-based Modeling Environment (DOME)	2	
	Effective Enterprise Learning	1	
	Implementation Dynamics (ID)	6	
	Incentives and Boundaries (IB)	3	
	Information Flow Modeling (IFM)	3	
	Other CIPD Research	10	
	Platform Architectures (PA)	4	
	Virtual Customer (VC)	4	
Center for Technology, Policy, and Industrial Development (CTPID)	CTPID Archive	3	Series of Technical Reports
	Cooperative Mobility Program	3	
	Ford-MIT Alliance	10	
	International Motor Vehicle Program	137	
	Labor Aerospace Research Agenda	10	
	Lean Aerospace Initiative	17	
	Lean Sustainment Initiative	3	
	Materials Systems Laboratory	1	
	Program on Internet and Telecoms Convergence	66	
	Program on Science, Technology, and Environmental Policy	5	
	Technology and Law Program	34	
Department of Ocean Engineering	Design Project Reports	7	
	Ocean Engineering Collection	1	
Laboratory for Information and Decision Systems (LIDS)	LIDS Technical Reports	1035	Technical reports
MIT Press	MIT Press Out of Print Books	74	Out-of print MIT Press books - access restricted to MIT
Singapore-MIT Alliance (SMA)	Advanced Materials for Micro- and Nano-Systems (AMMNS)	86	
	Computer Science (CS)	56	
	High Performance Computation for Engineered Systems (HPCES)	76	
	Innovation in Manufacturing Systems and Technology (IMST)	70	
	Molecular Engineering of Biological and Chemical Systems (MEBCS)	54	
Sloan School of Management		1058	Working papers
TOTAL		**2936**	

Table 1: Communities and Collections at DSpace@MIT

3.1.2 DSpace@Cambridge

Cambridge University Library, in association with the University Computing Service, has formulated a major project to provide the University with an institutional digital repository, 'DSpace@Cambridge'. This repository will provide a home for the increasing amount of material that is being digitised from the University Library's own printed and manuscript collections. It also has the ability to capture, index, store, disseminate and preserve digital materials created in any part of the University. These will potentially include scholarly communications (articles and pre-prints), theses, technical reports, archives of departments and the University as a whole, and other textual material, together with different formats such as multimedia clips, interactive teaching programmes, data sets and databases.

DSpace@Cambridge will involve formal collaboration with the MIT Libraries. The project is funded by a grant of £1.7 million from the Cambridge-MIT Institute and is due for completion in July 2005.

In parallel with DSpace@Cambridge the Cambridge-MIT Institute is also funding a complementary project, LEADIRS (LEarning About Digital Institutional Repositories Seminars), to promote strategic planning for institutional repositories in the UK higher and further education sector. LEADIRS runs a series of professional seminars along with working materials which will be provided to senior managers of institutions in the United Kingdom that are currently planning for, or in the midst of, the implementation of an Institutional Repository.

3.1.3 Others DSpace projects

In addition to the DSpace Federation members listed above, DSpace is being used or evaluated at a large number of institutions including Glasgow (see DAEDALUS project below) and Ohio State.

3.2 EPrints.org

EPrints.org is a collection of self-archiving and open access projects based at Southampton. As well as developing the GNU (open source) version of EPrints software, EPrints.org runs the CiteBase (an experimental bibliographic/citations database) and OpCit (reference linking and citation analysis for open access) services.

EPrints is currently the most widely used software for IRs.

3.3 DARE

DARE (Digital Academic Repositories) is a collective initiative by the Dutch universities to make all their research results digitally accessible. It can be seen as a national level, albeit federally structured, repository. Its programme of research projects is broadly similar to the JISC FAIR programme. The project was awarded 2 million euros for the period 2003-2006 by the Dutch government. DARE will follow open, international standards to ensure interoperability, nationally and internationally. All participating institutions will adopt the same standards, while retaining their own responsibility in setting up and maintaining their own repositories. See van der Vaart (2002) for more information.

3.4 FAIR

The JISC Focus on Access to Institutional Resources (FAIR) programme comprises some 14 projects. Virtually all are relevant to the IR agenda but perhaps the most relevant are RoMEO, SHERPA, ePrintsUK, TARDis and DAEDALUS.

3.4.1 RoMEO

The RoMEO project concluded in July 2003. It was unique among the FAIR projects in addressing rights issues. The key findings were published in D-Lib in September (Gadd et al. 2003).

RoMEO surveyed 524 academic authors in 57 countries across a variety of disciplines. Some 60% thought they initially owned copyright in papers, though 32% admitted they didn't know. 50% said 71-100% of their papers were co-authored, thus creating scope for disagreement on self-archiving. More than 60% were happy for others to display, print, save, excerpt from, and give away their papers, so long as given attribution and quotes were verbatim. In fact, authors were prepared to grant more liberal terms for the use of their own papers than they actually expected to be available to themselves for the use of other papers.

The project made a detailed analysis of journal publishers' copyright assignment terms – a useful database of these is maintained on the RoMEO website[4]. Around 90% of publishers ask for copyright, 6% for exclusive and 4% for non-exclusive licences, and 75% of authors are asked to warrant that their work has not previously been published. The headline finding for the self-archiving movement was that 55% of publishers' agreements formally permitted authors to self-archive (preprints only 36%, or postprints only 2%, or both 17%), and that many of the 45% balance would permit it if asked.

[4] http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html though responsibility for its maintenance will pass to SHERPA shortly.

The project also surveyed OAI data and service providers, revealing a degree of ignorance and/or unconcern with rights issues. Only 25% of data providers had licence agreements with their depositing authors, and 50% either just trusted the depositors, or simply provided a general warning statement. Regarding metadata protection, 50% of data providers thought (incorrectly) that metadata records were facts and as such had no copyright. Also 68% believed that though there was database right in metadata collections, this was "implicitly waived" within the OAI community.

RoMEO argue that the best way of dealing with the important rights issues created by the open archives, and illustrated by their research, is to develop machine-readable metadata schemes, compatible with OAI-MHP, to describe ownership and usage rights in the article and in the metadata itself. They developed a set of rights expressions using Creative Commons licences and are also planning to develop Open Digital Rights Language Initiative (ODRL) versions (i.e. XML instances) of Creative Commons licences that would conform to the ODRL XML schema. This work is continuing through the formation of OAI/RoMEO Technical Committee, due to report in Spring 2004.

3.4.2 SHERPA

The SHERPA project (Securing a Hybrid Environment for Research Preservation and Access) has been set up to encourage change in the scholarly communication process by creating open-access institutional e-print repositories for the dissemination of research findings. The outcomes of the project will be advice on the building and maintenance of IRs, guidelines on IPR and copyright issues, and advocacy material to publicise an institution's repository (Hubbard 2003; Pinfield 2003).

SHERPA is hosted by the University of Nottingham (Project Director Stephen Pinfield) and is a 3-year project from Nov 2002. Its partners are Edinburgh, Glasgow, Oxford, Leeds, Sheffield, York, the British Library and the Arts and Humanities Data Service (AHDS), with more partners in the pipeline. It aims to set up OAI-compliant e-print repositories (using EPrints software) at each of the partner sites. The project aims are:

- to set up thirteen institutional open access e-print repositories which comply with the Open Archives Initiative (OAI) Protocol for Metadata Harvesting (OAI PMH) using EPrints.org software;
- to investigate key issues in creating, populating and maintaining e-print collections, including: Intellectual Property Rights (IPR), quality control, collection development policies, business models, scholarly communication cultures, and institutional strategies;
- to work with OAI Service Providers to achieve acceptable (technical, metadata and collection management) standards for the effective dissemination of the content;
- to investigate digital preservation of e-prints using the Open Archival Information System (OAIS) Reference Model (an ISO standard for the long-term preservation of digital information, initially developed by RLG);
- to disseminate lessons learned and provide advice to others wishing to set up similar services.

3.4.3 ePrints UK

ePrints UK is a FAIR project due for completion in July 2004. Its aim is to develop a national service provider repository of e-print records based at Bath (Martin 2003; Day 2003).

As well as just harvesting the metadata, the ePrints UK service will add value to it, by adding/validating authoritative author names (i.e. removing ambiguity between different possible versions of an author's name by substituting a single authoritative version),

automatically assigning subject classifications, and automatically parsing bibliographic references into structured, machine-readable forms (using the OpenURL standard, which allows metadata (such as bibliographic information) to be contained within a URL. OpenURLs are "resolved" by a server that typically takes account of the user's context, for instance by linking the user to the appropriate copy of the article.).

Martin (the project manager) makes the point that the success of ePrints UK will depend on whether or not there is actually any metadata to harvest, and clearly fears that the project will not be successful because of factors beyond its control. She cites the "familiar barriers" to development of e-print services as IPR and publisher concerns; fears about quality of pre-print material; and control of metadata standards.

Day makes a similar point and cites the potential impediments to success as:

- copyright;
- peer review and quality control;
- long-term preservation;
- role of journals in scholarly communication: journals currently fulfil multiple, essential roles in scientific communication;
- different motives for writing and publishing papers;
- differences between subjects; and
- diverse nature of research institutions.

3.4.4 TARDis

The TARDis project (Targeting Academic Research for Dissemination and Disclosure), run by Southampton University, is planning to develop a multidisciplinary institutional e-print archive and assess and evaluate the activity within a library-led infrastructure. It is designed to tackle head-on the major problem faced by IRs, namely the lack of participation by faculty: "TARDIS will investigate and report on strategies to overcome the technical, cultural and academic barriers, which currently restrict the development and particularly the acquisition of content of institutional e-Print archives. It will develop a working model of a multidisciplinary institutional archive." The project runs from August 2002 until July 2004.

3.4.5 DAEDALUS

DAEDALUS is a project concerned with the establishment of a range of OAI-PMH-compliant digital collections at the University of Glasgow. These will include e-prints (both published and peer reviewed academic papers, and pre-prints and grey literature), theses, resource-finding aids and institutional documents. It runs until July 2005.

3.5 Other IR projects

3.5.1 Caltech

Caltech's CODA (Collection of Open Digital Archives) repository was established in 2001. It currently consists of some 11 archives with a further six listed as in development. The system allows document counts and a wide range of access statistics to be viewed by any user. The following table lists the active archives, their numbers of records and the total number of document accesses since launch. The vast majority of the content appears to be grey literature: theses & dissertations, technical reports and conference presentations.

Archive	*Description*	*No. records*	*Accesses*
CaltechBOOK	Books by Caltech Authors	2	8837
CaltechCDSTR	Caltech Control and Dynamical Systems Technical Reports	28	308
CaltechCSTR	Caltech Computer Science Technical Reports	410	58,278
CaltechEERL	Caltech Earthquake Engineering Research Laboratory Technical Reports	293	176,334
CaltechETD	Caltech Electronic Theses and Dissertations	773	417,551
CaltechGALCITFM	Caltech Graduate Aeronautical Laboratories (Fluid Mechanics) Technical Reports	5	n/a
CaltechLESSGS	Caltech Large-Eddy Simulation and Subgrid-Scale Modeling for Turbulent Mixing and Reactive Flows	29	384
CaltechLIB	Caltech Library System Papers and Publications	22	5,462
CaltechOH	Caltech Oral Histories	27	49,614
CaltechPARADISE	Caltech Parallel and Distributed Systems Group	54	4,510
cav2001	Fourth International Symposium on Cavitation. Hosted by Caltech, June, 2001	110	257,468

Table 2: Caltech CODA archives

3.5.2 eScholarship

The California Digital Library (CDL) eScholarship Repository, announced in April 2002, illustrates the continuum between digital libraries broadly conceived and institutional repositories. The CDL launched the eScholarship repository, a web site and a suite of digital support services, to distribute academic research and working papers of University of California faculty. eScholarship uses the OAI metadata harvesting protocol to provide interoperability.

The CDL initiative includes a suite of digital services to store and disseminate faculty research in digital formats. The CDL system uses the web-based bepress (vendor) system to manage paper submission, processing, and dissemination. Additionally, the system also supports a topical alerting service that alerts users to new content in their specified areas of interest.

There are about 2450 e-prints (both pre-prints and post-prints), mostly in the social sciences, mainly economics and related areas (e.g. transportation). The stated policy is to accept "journals, peer-reviewed series, working papers, discussion papers series, and other electronic forms of scholarship". Most of the documents we viewed were working papers. Content is stored as PDF only (but the system will accept and convert Word, RTF etc. into PDF).

3.5.3 CARL

The CARL Institutional Repositories Pilot Project is an initiative to implement institutional repositories at several Canadian research libraries. The project, which is spearheaded by the Canadian Association of Research Libraries, was launched in September 2002 and has 12 libraries participating. The repositories will be searchable using one interface and freely accessible to anyone with an Internet connection. The ultimate vision is to have a number of robust and interoperable archives containing Canadian scholarly output that will form a part of a larger global system of repositories.

In the initial phase of the project, participants are sharing best practices and lessons learned in order to assess the feasibility of IRs in the Canadian context. Members are experimenting on a trial basis with a variety of software types (EPrints, DSpace and bepress's eScholarship), content, and archiving policies, among other things. To date, four of the participants have their IRs up and running, while the others are at various stages of planning or implementation.

CARL has published a Position Statement on Institutional Repositories (CARL 2003). It sees the benefits of (Canadian) IRs as:

- increasing the visibility of Canadian researchers and institutions;
- increasing the accessibility and impact of Canadian research domestically and internationally;
- long-term preservation of research output of Canadian academic institutions;
- increasing the proportion amount and diversity of scholarly output that is collected and preserved (cf. traditional collections policies focussed on published materials);
- facilitating more timely access to research and scholarship.

3.5.4 Ohio State University (OSU) Knowledge Bank

A project team under Joe Branin, Director of Libraries, has developed a detailed proposal for an IR at OSU to be known as the OSU Knowledge Bank. The proposal was submitted in June 2002 and is freely available from the Knowledge Bank website (Branin 2002). The proposal differs from other IR descriptions in that there is a "knowledge management" theme: amongst other things, the IR is seen as a way for the university to help manage its intellectual content – "an important and valuable asset of the University".

At present there is a very basic DSpace implementation that has yet to be populated.

3.5.5 Utrecht

DISPUTE, the institutional repository of the University of Utrecht was originally scheduled for release at the end of 2002. The site is currently (in early 2004) in a preliminary state and states variously that the final site will be available in October 2003 (sic) or by the end of 2003 (sic). All this suggests the going has not been as easy as was originally envisaged. At present there appears to be about 400-500 maths and 40-50 physics e-prints, plus some other documents and theses. Several of the e-prints we viewed were scanned from published literature (e.g. *Physical Review*). The site also appears more a portal to a disaggregated collection of resources than a single IR.

3.5.6 ARNO

The Academic Research in the Netherlands Online (ARNO) project, run from September 2000 until September 2002, sought to design and implement university digital archive servers to preserve the academic output (including research reports, pre-prints, theses and dissertations, and articles published in regular scholarly journals) of member institutions. The project's goal was to make the repository freely accessible via OAI interoperability standards. The project was being implemented by the library staffs of the University of Twente, the University of Amsterdam, and Tilburg University.

Specific project goals included:

- Connecting the document servers to international distributed digital archives and to the Dutch national information infrastructure;
- Developing an infrastructure that will couple with the production processes of scientific publishers and offer a good basis for handling peer review.
- Connecting seamlessly to digital learning environments.

The ARNO software is available via Open Source licensing.

DARE (see section 3.3) has effectively superseded this project.

3.5.7 Max Planck Institutes

The Max Planck Gesellschaft group of research institutes in Germany has built its own repository software, eDoc Server. MPG recognises that the motivations for introducing and using the eDoc Server will differ from one discipline to another. From an institutional viewpoint the eDoc Server aims to:

- "build a comprehensive resource of scientific information produced by the Max Planck institutes, providing a stable location for its preservation and dissemination;
- "increase the visibility of the intellectual output of the Max Planck Institutes in all the forms it takes in the era of the Internet;
- "strengthen the Society and the scientific community in negotiations with publishers about the ownership of scientific research documents at a time where sky-rocketing journal prices and restrictive copyright undermine their wide dissemination and persistent accessibility;
- "contribute to a world-wide, emerging scholarly communication system, which exploits the full potential of the Internet and the digital representation and processing of scientific information."

The MPG eDoc server is currently live at many MPG institutes. It contains full text articles of conference papers and e-prints and links to commercial publishers' online journal services for published articles.

4 Issues for IRs

In this section we briefly discuss the main issues facing those establishing IRs. (The issues that IRs create for publishers are discussed separately in section 9.2 below.)

4.1 Cultural issues affecting faculty take-up

The biggest problem facing those setting up IRs is persuading faculty to use them. Outside a few disciplines (e.g. physics, computer science and economics) there is little tradition of preprints or working papers and apparently still little interest in self-archiving. Academics may be radical in their thought but they are conservative in their behaviour, and there is a great deal of inertia in the current publishing systems. The point is made repeatedly in all the literature that organisers of IRs need to budget time and money for advocacy (marketing campaigns, meetings, flyers, websites, emails, etc.), for training users and producing guidelines, and for ensuring the interface design does not put off potential users.

The data quoted in this report shows that take-up rates for IRs have to date been very patchy, especially where the deposit of materials depends on the decision by individuals to self-archive their material. (The exceptions are theses and dissertations, where it is feasible to compel postgraduate students to submit electronically, and collections of technical reports or working papers held at a departmental or institutional level.) The rate of deposit of new records typically falls off sharply after an initial burst (see section 8.3 below).

It is argued that *institutional* repositories will have both the prestige and the clout with faculty to persuade them to start self-archiving. The SPARC position paper, for instance, argues that both parties – the institution and the faculty – stand to gain in different ways from the establishment of IRs: the institution from increased visibility of its output and the resulting prestige, the faculty member from the greater visibility of their own work and the resulting research impact. There is little evidence yet from the data reviewed in this research to support this: where faculty are using IRs to self-archive they are overwhelmingly from the disciplines that are already self-archiving (physics, economics, etc.).

4.2 Intellectual Property Rights/Copyright

Institutional repositories raise issues of rights - not just copyright but other intellectual property rights such as patent rights. The copyright issues have been extensively reviewed by the RoMEO project, described elsewhere in this report. As RoMEO notes, there are rights associated with metadata as well as with the content itself.

4.3 Existing digital collections

New IRs at most organisations will have to take account of existing collections. Some of these (e.g. personal self-archiving websites) may be best subsumed within the IR; others (e.g. collections of theses or library/departmental collections of digital objects such as digitised images) might be better left separate but interoperable, at least during the early phases.

4.4 Organisation and administration

Who will manage the IR? Where will it be located? What will be the relationship between the centre and the departments? DSpace, for instance, is designed to allow individual "Communities" to run their own "Collections" in their own ways.

4.5 Funding/business model

How will the IR be funded, and what business model will it adopt? This issue is discussed below, see sections 6.3 and 6.4.

4.6 Preservation

The issues raised by the long-term preservation of digital objects are very far from solved. There are at least four approaches to preservation (migration; technology preservation; emulation; and persistent object preservation) and there is significant research going on under nine headings (see Smith 2003, p.46). This is an uncertain area with an uncertain future price-tag. IRs can attempt to limit future liabilities (or disappointments) by restricting the number of digital formats that will be fully preserved, as DSpace has done.

4.7 Accession Policies

The IR has to decide:

- who is allowed to deposit materials;
- what types of materials can be deposited (e.g. pre-prints, post-prints, working papers, theses, chapters, datasets, etc.);
- what digital formats will be accepted (and which formats will be migrated);
- quality assurance procedures (e.g. none, or approval by departmental head, or even peer review);
- storage limits etc.

4.8 Open access

Although IRs can be seen as part of the open access movement, this is not necessarily the case. IRs may well want to restrict access to part of their content to their own network only. Many of the IRs reviewed as part of this research had at least some content restricted to campus access: for example, MIT's DSpace limits access to its collection of out-of print MIT Press books, and several sites restrict access to student theses.

4.9 Central vs. institutional repositories

This argument is somewhat theological. Central, subject-based repositories have been very successful in a few fields (physics, computer science, economics, cognitive science) but have failed so far to migrate to other disciplines. Institutional repositories may offer a new route to reduce the inertia among the non-archiving communities, as discussed elsewhere.

From the perspective of a web-based *searcher* (i.e. reader), however, there is (in theory) no difference between the two repositories, as articles are discovered on the basis of metadata independent of location (author, title or whatever), collected via the same OAI protocol. (The searcher would be using an interface on a third party site, i.e. an OAI service provider.) In practice, of course, there are significant benefits in centralised systems, for instance in terms of standardisation, single metadata schemes, etc. As PubMed Central (the US National Library of Medicine's digital archive of life science literature) put it:

> *"PubMed Central, by storing data from diverse sources in a single repository with a common format, makes the data more accessible and easier to use and opens the door to greater integration with related resources, such as the variety of databases available in NCBI's Entrez system. Full text can be searched and relevant material located efficiently, regardless of its source. Storing all articles in a uniform and well defined (tagged) structure allows other features, e.g., searches focused on the Methods section*

of articles, or links from the literature to existing resources such as sequence databases and structure viewers, to be applied consistently across the entire collection. It also enables the development of tools to further integrate the literature with the many information resources available to scientists, clinicians and others. In addition, PMC presents material to the user in a uniform style while still clearly maintaining the identity of each journal."

4.10 Metadata

Without at least some metadata, nothing can be found. But too comprehensive a metadata requirement will inhibit users. DSpace has come down in favour of a minimal set of three compulsory fields, while allowing (and encouraging) many more.

Decisions also have to be taken on which metadata schemes to use[5]. Metadata requirements may vary from field to field and DSpace allows its separate Communities to use their own discipline-specific metadata schemes.

4.11 Technological

The main technological issues for a new IR will be to select from the several viable software options available (see above) and to size the server and storage space requirements (see e.g. Branin 2002; Barton and Walker 2002). The key technical requirements are to do with:

- interoperability: This is generally achieved via OAI-compliance;
- persistent identifiers: DSpace uses the CNRI Handle system.

[5] For example, see the 16-odd metadata schemes listed on the Open Archives Forum website at http://www.oaforum.org/oaf_db/list_db/list_metadata.php

5 Uses

The main documented uses for IRs are discussed below.

5.1 Scholarly communication

This is of course a very broad category but it is the main driver for IRs. There are currently two main thrusts: first, to reform scholarly publishing in various ways and secondly to support the new digital scholarship (for more on the latter see Smith 2003).

It is not clear however why *institutional* repositories should be more effective in reforming publishing than other types of repository (e.g. subject-based). The main argument seems to be that because institutions stand to gain from IRs (e.g. additional prestige through making their research output more visible), they will exert influence or direct or indirect pressure on faculty to self-archive their material. We have not seen any evidence for this happening to date, other than in the compulsory electronic deposition of postgraduate theses and dissertations.

5.2 Education

Repositories can support education not only by supporting campus teaching and "digitally captur[ing] and preserv[ing] many of the events of campus life—symposia, performances, lectures", but also by globally disseminating teaching via the web (e.g. MIT's OpenCourseWare initiative, which aims to collect and make available MIT course materials).

Repositories are currently being used to store course materials created by faculty. IR software is not however designed to act as a virtual learning environment, so it seems likely that distance learning *per se* will continue to be delivered via more specialist applications.

5.3 e-Publishing

The development of new publications or publishing models based on IRs. For instance, Dennis Hall (the associate provost for research and graduate education at Vanderbilt University) has suggested that IRs may become the new university presses (Hall 2003). He argued that many more universities will experiment with IRs and that it will be a natural step for local peer-review quality control measures to evolve into broader *de facto* publishing ventures. But he sees this as a long-term process, with universities becoming competitors to existing commercial and not-for-profit publishers over the next one to two decades.

Meanwhile, the eScholarship repository at the University of California offers software tools to create online journals. Such journals have to be sponsored by a UC research centre or department and be freely available online. *Dermatology Online Journal* was the first journal to migrate to the eScholarship system during 2003. It has been followed by others including the launch of the *San Francisco Estuary & Waterway Science* in October 2003 and a peer-reviewed series of articles and monographs form UC International and Area Studies in conjunction with the UC Press.

5.4 Collection management

Digital objects on university networks are currently largely uncatalogued, widely scattered and not managed. There is no central catalogue or database for such materials. Such materials, which may be on personal or departmental websites, are therefore difficult to discover and use, and difficult to keep track of and preserve. Migrating such collections to a central IR would address these issues.

5.5 Long-term preservation

This is seen as a key function of an IR for most participants and commentators. There is less certainty about how it will be achieved, and whether indeed IRs are the best placed to manage the issues involved. Lynch argues that only research institutions have the perspective and interest necessary to commit to the long term.

The challenges are potentially substantial; the US Congress awarded the Library of Congress around $100m to build a programme to address digital preservation. However some in the open access movement, focussed on widening access to research articles, have argued that concerns about preservation are over-stated and of a lower priority than accelerating the spread of self-archiving.

The costs associated with long-term preservation are discussed in more detail below, see section 6.3.4.

5.6 Institutional prestige

It is argued (e.g. Crow 2002) that by collecting together and making easily visible the university's research output, its prestige may be enhanced. This issue was discussed above, see section 2.2. The section on Research assessment exercises below is also relevant.

5.7 Knowledge management

IRs may offer university management a tool to manage IPR more effectively. This is likely to be a potentially controversial application, because it may be seen as being concerned not just with *managing* IPR, but also to do with gaining *control* or *ownership* which many academics would resist.

Nonetheless, the Ohio State Knowledge Bank proposal mentioned above (3.5.4) talks about managing the university's intellectual content, "an important and valuable asset of the University". This kind of knowledge management approach would have more relevance to patent-type IPR than to copyright.

5.8 Research assessment exercises

Harnad (2001) for instance has argued that research assessment exercises could be conducted far more efficiently if the research outputs of each institution being assessed were easily visible through a web repository (ideally supported by open citation analysis (i.e. a non-proprietary impact factor analysis which might include new kinds of online citation as well as references in peer-reviewed text) and a standardised online CV for research staff, with links to full text versions of the publications listed). In the UK, JISC and RAE (the body that conducts research assessment exercises on behalf of the higher education funding bodies) may support this.

6 Management

This section looks at how established repositories are managed within institutions, and who is responsible for their maintenance and providing access to them, and also gives some data on the costs of establishing and maintaining an IR.

6.1 Organisation

It seems likely that all substantial IRs based at large, diverse organisations (such as research universities) will adopt some kind of delegated or federal structure. For example, DSpace is designed to be organised into "communities" and "collections". Similarly the Caltech CODA site is organised into a number of different departmental-based archives, and Ohio's Knowledge Bank is planned to be highly federal. There are good reasons for this. First, no single accessions policy (i.e. who is allowed to deposit what, with what QA/peer review process) is going to suit all disciplines and departments – if nothing else, different disciplines are likely to move at different paces. Secondly, the types of document (and/or their nomenclature) will vary by discipline (e.g. science preprints, economics working papers, engineering technical reports, humanities monographs). Thirdly, different disciplines will want to use different metadata schemes. Fourthly, different disciplines will have different requirements for document formats and sizes.

6.2 Administration and maintenance

As mentioned below, the very large majority of IR sites studied in this research were run by the library (or information services) at the host institutions. Exceptions were run by central (national) organisations such as CCSD (Centre pour la Communication Scientifique Directe) in France or the National Centre for Science Information in India. One IR (Hofstra) was under the management of the Office of the Provost.

Projects to develop IRs typically involve a three-way partnership between the library, the IT function and the bursar's/provost's office. (For example see the MIT case study (MIT 2003) or the Ohio Knowledge Bank proposal (Branin 2002).) There are a number of reasons cited as to why the library is the appropriate locus for the leadership of such projects:

- librarians have a greater and/or broader awareness of the pressures on library budgets and the impacts on content availability;
- the library is a natural home for the stewardship and long-term preservation aspects of IRs;
- they also tend to have greater awareness of the self-archiving movement;
- librarians have a greater awareness of the issues involved in digital preservation;
- librarians possess relevant skills that are typically not found elsewhere on campus, for example metadata creation and maintenance;
- participation in existing digital library projects has given them awareness of other issues such as persistent identifiers (see section 2.5.1), appropriate copy (see section 9.2.2), and so on.

The MIT case study however makes it clear that partnership is essential both with IT (to develop and maintain the systems) and with senior management (to gain the clout to secure start-up budgets and to gain buy-in from the faculty at large).

6.3 Costs

6.3.1 Start-up costs

DSpace development at MIT was supported by a $1.8m grant from Hewlett Packard, plus 3 FTE HP staff on secondment, plus $400,000k in system equipment –a total of say $2.4-2.5m. This cost is treated as written off in MIT's business plan for DSpace (Barton and Walker, 2002).

The DSpace software has now been released under an open source licence, so for other institutions wanting to set up a new IR using DSpace (or EPrints) software the technology start-up costs are limited to a server plus backup plus installation and configuration time. This is estimated to cost between $10k and $50k. A simple EPrints server could be set up for even less.

6.3.2 Running costs

The main incremental costs are server costs and people. MIT's business plan for DSpace estimates annual running costs of about $285k, broken down as follows:

Staff	$225k
Operating costs	$ 25k
Systems equipment	$ 35k

These costs certainly do not look to be overestimates of the true costs of running such a complex system, though the figures that eventually appear in university accounts may be less if staff costs are hidden within other (e.g. departmental) budgets.

6.3.3 Activities

The activities driving the running costs of an IR are (Barton and Walker 2002):

Management

Advocacy

End-User and Community-level support/help

Training

Community management

Review and approval processes (these activities are most likely to be borne by end-users and/or their departments rather than the central IR team)

Metadata development, creation and maintenance

Digitisation of hard-copy materials

File format conversion

Systems management (systems monitoring, testing & debugging, system administration, etc.)

Backup and recovery

Technology upgrade, development, research

And no doubt others, such as attending conferences on IRs and miscellaneous office expenses.

6.3.4 Long-term preservation costs

One of the likely largest costs over the long term will be the costs associated with long-term preservation, for instance data migration and conversion. However they are not only among the largest costs, but also by far the least known and indeed least knowable. So a commitment to host an IR amounts to an implicit commitment to an unknown amount of work at some point in the future and as Bill Hubbard says "there is now, quite rightly, a general reluctance to fund institutional scale projects without a clear and well-thought analysis of future implications" (Hubbard 2003). Proponents of IRs (Hubbard among them) argue, however, that it would be a mistake to hold back from development because of this uncertainty, at least in part because "the important issues and trickier questions will only emerge through the population and use of [IRs]".

One strategy for mitigating the potential future impact of data is to commit to migrating only a subset of available digital formats, typically the open non-proprietary standard formats. For instance, DSpace at MIT will "accept digital objects in all formats, including individualized formats. The policy further defines preservation service levels, categorizing file formats according to whether they are known and supported, known and unsupported, or unknown and unsupported. For 'unsupported' file formats, DSpace merely guarantees to preserve and return the bits, while for 'supported' formats DSpace makes a further promise to maintain the usability of the content in its original context."

6.4 Business models

The MIT DSpace business model assumes that funding will come from MIT itself, from corporate and federation partners, and from revenue from premium services (see below).

MIT DSpace will offer core services free of charge. These include submission, hosting, access services (including notifications), storage and preservation management services necessary to ensure the longevity of deposited materials, community management services (consultative services, guidance, template policies, etc.); community support services (web and telephone support); and system management.

Premium services will be charged for. These include e-conversion services (digitisation and digital format conversion); metadata services (consultancy, metadata research, metadata creation and support); custom repository services (e.g. for communities whose storage requirements exceed the norm); and user reporting services (research alerts, targeted notification services, hot topic citations, custom reporting services). MIT see the charges for premium services as not being primarily to generate funds but "to allow MIT to offer as complete and valuable a service as possible to users while controlling the impact on the Libraries' scarce resources".

Another possibility is that larger, better resourced and more experienced institutions may offer IR services to other universities on a local, regional, national or other basis. This would allow some of the total costs to be centralised, probably offering economies of scale.

7 Requirements for deposit of materials

There is a great deal of variation between institutional repositories in their requirements for deposit of materials. Furthermore many IRs are still only at an experimental or start-up phase and their policies in this regard have yet to be finalised. With these caveats, this section looks at what the IRs on our list are currently doing.

7.1 Allowed users

Operational IRs generally only allow members of the institution to deposit documents. Some IRs encourage individual members to deposit, others prefer a moderated process (e.g. through the department).

There is a debate as to whether to allow undergraduates to deposit materials; where it is allowed it is likely to be subject to a degree of supervision and control.

(Some of the more experimental IRs (e.g. Glasgow, ANU) allow any user (including this author) to register and then upload documents, though these are then presumably filtered out in the post-upload approval processes.)

7.2 Types of document

As shown below, the aggregated content currently posted on the IRs covered appeared to divide into:

- 22% e-prints (both pre-prints and post-prints);
- 20% theses and dissertations;
- 58% other documents (grey literature (e.g. technical reports and working papers), and this also includes a collection of 14,000 digital images at Virginia Tech).

The eprints appeared to be mainly preprints but this was not accurately quantified.

Documents are mainly text-based articles of various types; there is currently little evidence of more complex digital materials, datasets etc. though some IRs state that they are planning to allow these at a later stage.

Some IRs have a policy of not permitting published material to be deposited.

7.3 Submissions approval processes

It is common for deposited materials to be subject to a check before release. For example, this extract from the Caltech CODA policy for the caltechCSTR (Computer Science Technical Reports) archive:

> *"CS Department Faculty approves content and inclusion in the caltechCSTR digital archive. The repository managers review metadata and contents prior to final deposit into the archive."*

It seems likely that most such reviews cover quality issues such as completeness, appropriateness of content (e.g. type of document), metadata, etc. rather than the quality of the content itself.

7.4 Copyright

Most IRs require the depositing author to grant (non-exclusive) electronic rights to the IR and to warrant their right to do so. For example, under the Caltech CODA agreement, the author will:

"grant to the California Institute of Technology (Caltech) the irrevocable, non-exclusive royalty free right to reproduce, distribute, display, and perform this work in any format including electronic formats throughout the world for educational, research and scientific non-profit uses during the full term of copyright including renewals and extensions via the Digital Collections mechanisms maintained by the Caltech Library System. I also hereby grant to Caltech the nonexclusive right to sub-license these rights to others should the Institute forego the ability to maintain distribution. I warrant that I have the copyright to make this grant to Caltech unencumbered and complete."

7.5 Formats

There is a wide spectrum of policies on allowable formats, from some accepting only PDF (e.g. Nottingham) at one end to accepting any digital format (e.g. DSpace at MIT) at the other. All IRs accepted PDF. Those only accepting PDF often offered tools to convert other formats (e.g. Word, LaTeX) into PDF.

Other formats accepted included HTML, XML, Postscript, RTF, MPEG, TIFF, ASCII, DVI, GIF, LaTeX and others.

DSpace, while accepting any format (including individualised formats), will only support a limited number of non-proprietary formats (e.g. XML, TIFF). Supported formats will be migrated as necessary to newer standards to ensure they can continue to be used. Non-supported formats will have their "bitstream" preserved but there is no guarantee that the file will be readable. Proprietary formats such as Microsoft Word are unsupported, (perhaps surprisingly) PDF.

7.6 Compulsory deposit

Several institutions have made it compulsory for postgraduate students to deposit their theses and dissertations electronically. In some cases, the student can select an option for the thesis to be viewable only from the campus network, in other cases there may be a general decision to restrict historical collections of theses to the campus, but in general new theses are freely available.

Other than postgraduate students it is rare for institutions to have policies for compulsory deposit. The Queensland University of Technology, however, recently introduced a policy whereby research output – except where "commercialisation or individual royalty payment or revenue for the author or QUT" is expected – must be deposited in the QUT repository. The policy statement notes that "In effect it applies to the corpus of refereed research literature, conference proceedings, and other non-refereed output" of QUT faculty[6].

7.7 Removal of papers

IRs generally discourage or forbid the removal of documents once deposited. For example, at the eScholarship repository, authors may request that a paper, or a version of a paper, be removed but even when the paper is removed a citation to it will still remain. Other policies do not permit removal other than in exceptional circumstances.

7.8 Example policy statement

The University of California eScholarship policy statement is attached as an example in the Appendix (section 1.1).

---

[6] See http://www.qut.edu.au/admin/mopp/F/F_01_03.html

8 Quantification of repositories and records

In order to quantify the scale of IRs and the types of content, we look first at the wider set of about 250 OAI repositories (of which IRs are a subset), then at a list of 45 IRs developed for this research. Data on the 133+ repositories using EPrints software (which overlap with IRs) are then reviewed and finally we look at some data on academics' self-archiving behaviour at Edinburgh.

8.1 OAI repositories (data providers)

OAI data providers are websites that allow their metadata to be harvested under the OAI-MHP protocol. Institutional repositories are generally OAI-compliant, and hence are OAI data providers; however not all data providers are IRs (for example, the largest subject-based repository, arXiv). *Figure 1* shows how the number of OAI repositories (excluding OCLC thesis data) and the mean number of records per repository has grown from 1999 to the beginning of 2003 (ECS 2003):

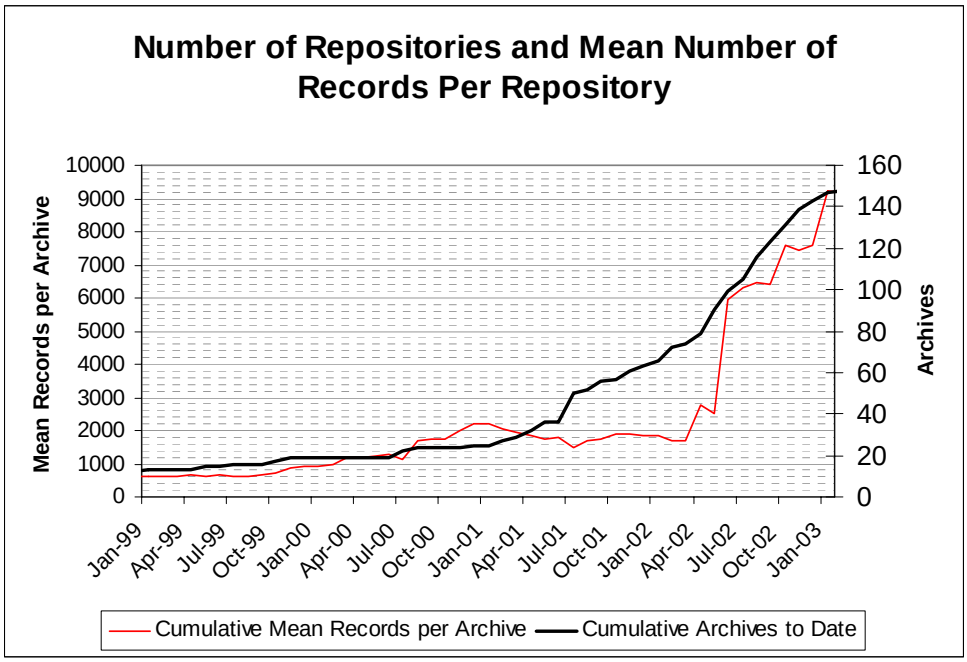

Figure 1: growth of OAI repositories and records

To bring this up to date, as of 4 December 2003, the OAI service provider OAIster was "…serving 2,228,430 records from 243 institutions". The number of records appears to underestimate the actual number held by some sites, e.g. RePEc (Research Papers in Economics) has about 210,000 records but only 53,003 are visible via OAIster.

The number of records per site ranged from 1 to 265,462 (CiteBase); a frequency distribution is shown below in *Figure 2*.

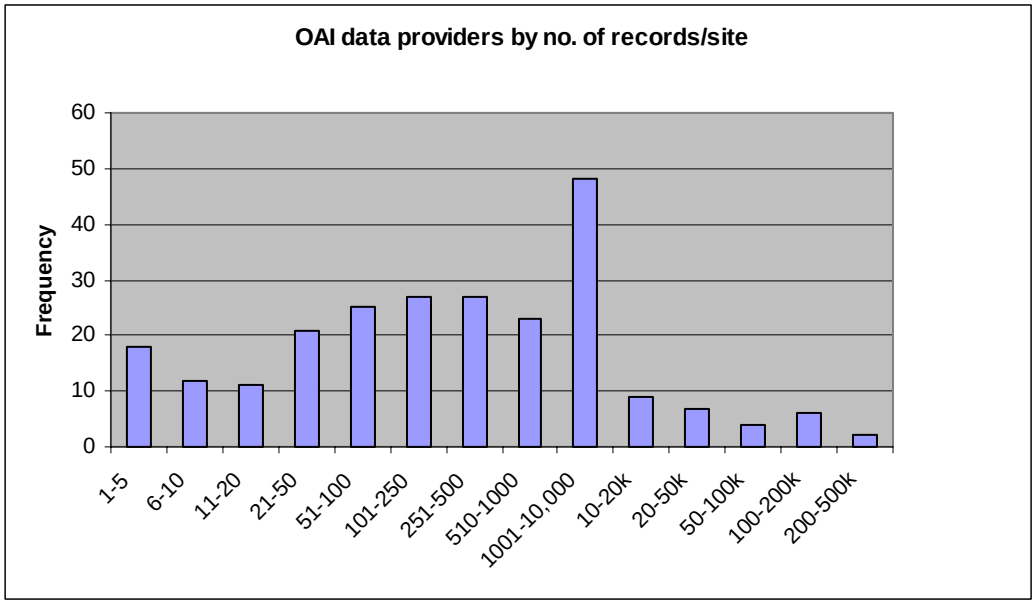

Figure 2: Distribution of sizes of OAI data providers

The median number of records per site is 314. More than a third of sites provide fewer than 100 records, and more than two-thirds provide fewer than 1000. (Whether these sites are "small" or large" depends on the type of records – 500 technical reports in a narrow field could be a very large archive, whereas 500 journal articles would be a small fraction of the output in most fields.)

The data providers represent a wide variety of organisations and offer a wide variety of types of material. Not all these records represent unique digital objects, e.g. the largest, CiteBase, is a bibliographic database rather than a repository. A full analysis of the types of content represented by these 243 sites is beyond the scope of this report, but by inspection of the larger sites it appears that peer-reviewed published literature in its final form makes up only a small proportion of sites but a fairly substantial part of the total number of records:

- theses and dissertations, e.g. Digital Library of MIT Theses (8676 records), Virginia Tech (4669 records);
- grey literature, e.g. NASA Langley Technical Reports (4085 records), NASA Aeronautics Reports (7640 records);
- digital library collections, e.g. Library of Congress American Memory Project (145,501 records), LOUISiana Digital Library (15,706 records);
- e-print archives, e.g. arXiv (253,205 records), RePEc (53,003 records);
- central library databases, e.g. PubMed Central (126,301 records);
- commercial publishers, e.g. BioMed Central (10,108 records), Institute of Physics (179,212 records).

8.2 Institutional repositories per se

Using the SPARC website's list of IRs as a starting point and adding additional sites that were discovered during this research has produced a list of some 45 IRs. This list is likely to be incomplete or inaccurate, not least because of problems of definition. These IRs were inspected and information gathered under these headings: software, content, features, formats,

subject areas, numbers of documents (theses, e-prints, others), launch date, contact details, managing organisation/dept, and miscellaneous notes.

Excluding the CERN Scientific Information Service site because of its scale and subject-specific nature, the total number of documents on these 45 sites was about 42,700[7]. The mean number of documents per site was 1250 and the median number 290. Broadly speaking, the content appeared to divide into:

- 22% e-prints (pre- and post-prints; we are not able to give a precise split between pre- and post-prints but post-prints appeared to be much rarer than pre-prints)
- 20% theses and dissertations;
- 58% other documents (including grey literature (e.g. technical reports and working papers), and a collection of 14,000 digital images at Virginia Tech. Datasets did not appear to be common – none was viewed in this review.)

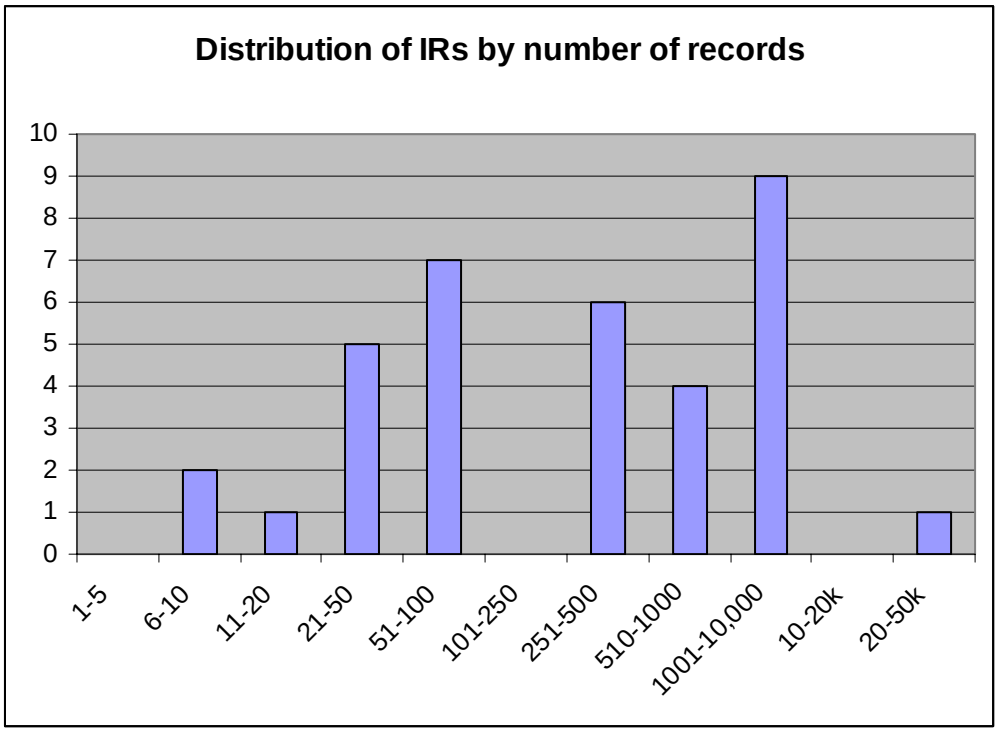

Figure 3: Distribution of IRs by number of records

The distribution of sizes of IRs is shown in *Figure 3*. By comparison with *Figure 2* it can be seen that IRs generally have somewhat fewer records than OAI data providers. This probably reflects their early stage of development.

The main subject areas covered were physics, mathematics, computer science and economics but there was also a tail of other subjects including linguistics, philosophy and some humanities. One notable omission was medicine and the clinical sciences, probably reflecting the lack of preprint culture in these disciplines, which is generally ascribed to the importance of peer review where scientific information is used by non-scientists (i.e. clinicians) and where errors may have serious consequences. Chemistry was only mentioned on one site reviewed, again presumably reflecting its lack of preprint culture.

[7] The CERN Data Service holds 600,000 records including 250,000 full-text documents

Where possible to determine, the very large majority of sites were run by the library (or information services) at the host institutions. Exceptions were run by central (national) organisations such as CCSD (Centre pour la Communication Scientifique Directe) in France or the National Centre for Science Information in India. One IR (Hofstra) was under the management of the Office of the Provost.

8.3 EPrints

EPrints version 1 was released on an open-source licence in January 2001 and the number of sites using it has grown from the 2 or 3 preceding the release to 125 at the time of writing. The following data (*Figures 4 and 5*) are about 10 months out of date but are the most current available from the EPrints.org site:

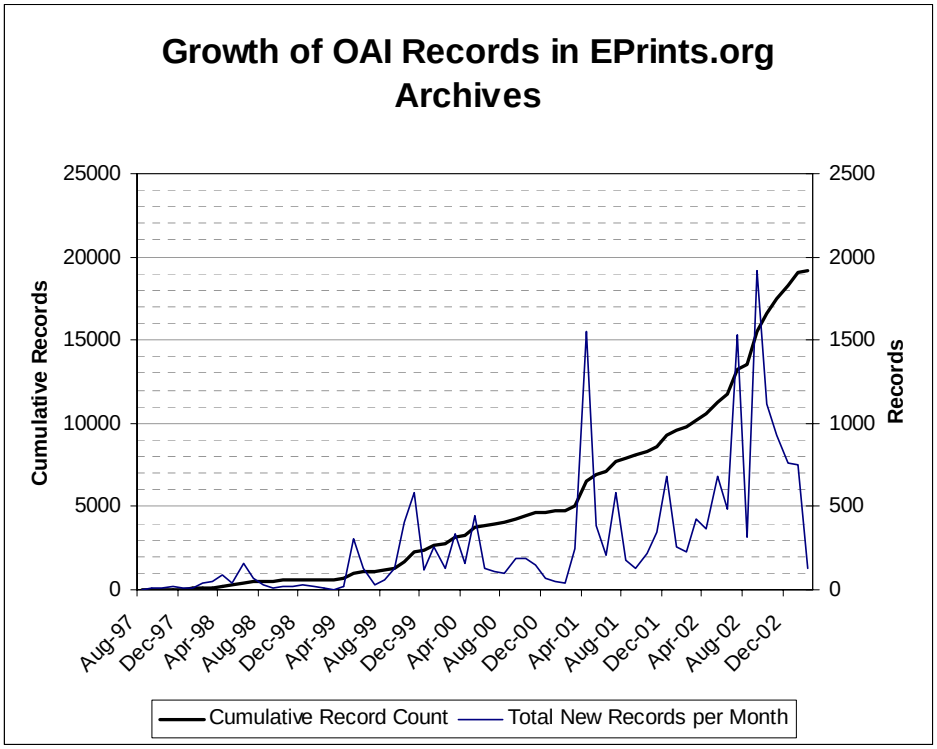

Figure 4: Growth of OAI records in EPrints-based archives

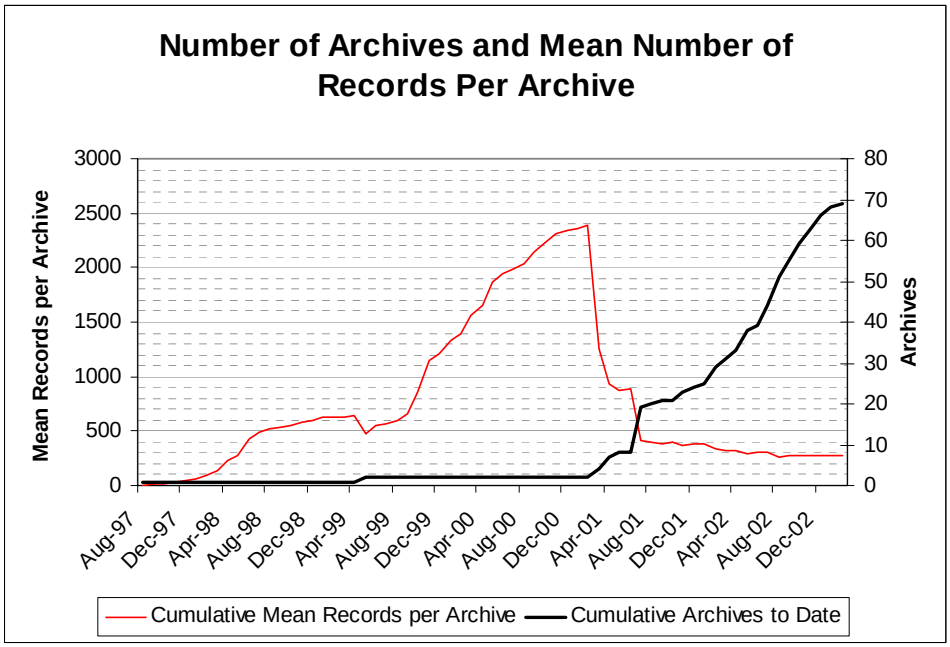

Figure 5: Number and average size of EPrints-based archives

The average (mean) number of records per archive has fallen sharply corresponding with the launch of new archives in 2001 and 2002, and stood at 277 at the beginning of 2003. If the original three archives (Cogprints and ECS at Southampton, and Lund University's eprint server) are excluded, the average number of records per archive was 130.

The difficulty that new repositories have faced is illustrated by *Figure 6*, which shows that the numbers of new records added falls sharply after the first couple of months, possibly reflecting the lack of sustained demand for the service once the initial promotional activity around the launch has dried up.

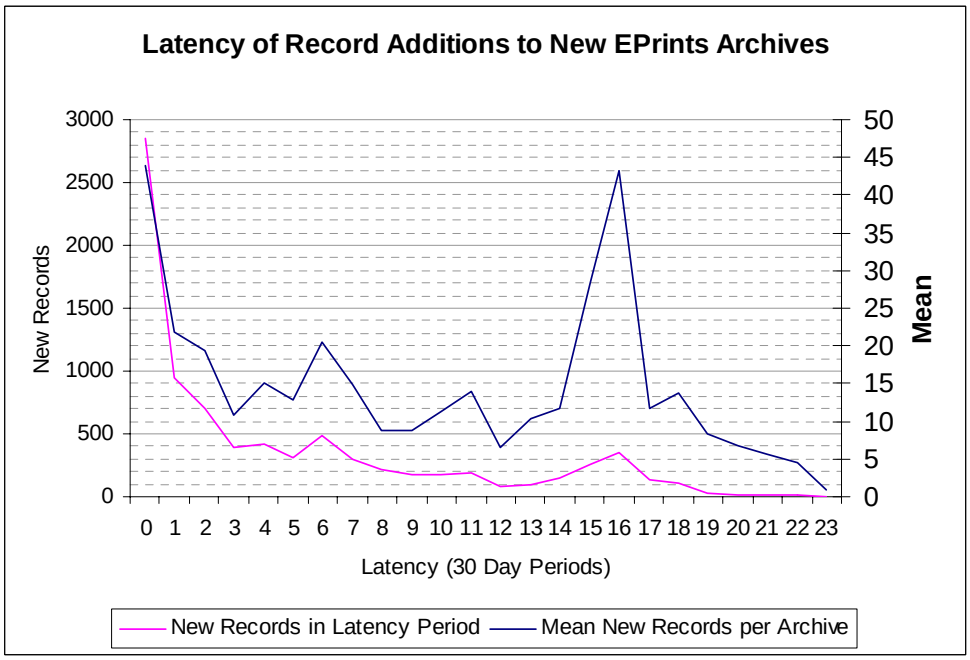

Figure 6: Latency of record additions to new EPrint-based archives

8.4 University of Edinburgh data

As part of the SHERPA/Theses Alive! Projects, Theo Andrew at the University of Edinburgh studied the material posted by faculty onto personal/departmental websites in ed.ac.uk domain (i.e. within the University of Edinburgh network) and compared the results by discipline (Andrew 2003).

Over 1000 papers were found. The problem was seen to be that they were very widely dispersed and hence not easily found (these sites did not of course use OAI-MHP). Also, personal web sites tended to be ephemeral, so long-term preservation was very uncertain.

A surprising (to Andrew) finding was "the relatively low volume of preprints found on personal Web pages. This could perhaps be related to the success of e-print repositories. Another significant factor is that most papers or theses found online were part of a researcher's publication list in his or her online CV, which essentially showcases research interests and credentials. Preprints do not have anywhere near the same impact factor as those papers from accredited journal titles, so it is possible that researchers would favour only putting their most impressive work in their online CV." (Here Andrew is using "papers" in contrast to pre-prints, i.e. to mean post-prints.)

The results confirmed big differences between disciplines, e.g. from the 33% of faculty members in the School of Informatics who were self-archiving, down to zero in some departments. Andrew concluded that there was a direct correlation between willingness to self-archive and the existence of subject-based repositories. Some departments did not self-archive because of concerns about legality under copyright laws. A low volume of preprints was found on personal web pages (as opposed to departmental).

The following graphs, taken from Andrew's paper, highlighted the differences in self-archiving behaviours between Science and Engineering, Humanities and Social Sciences and Medical/Veterinary Medicine.

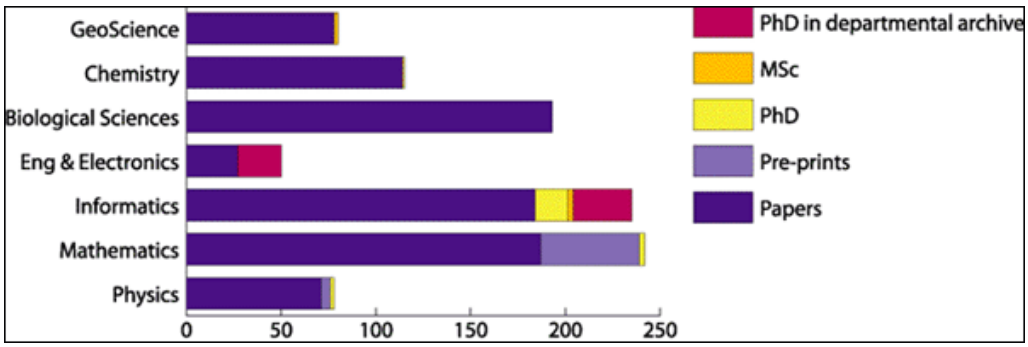

Figure 7: Volume and type of research material presently available in the S&E ed.ac.uk domain

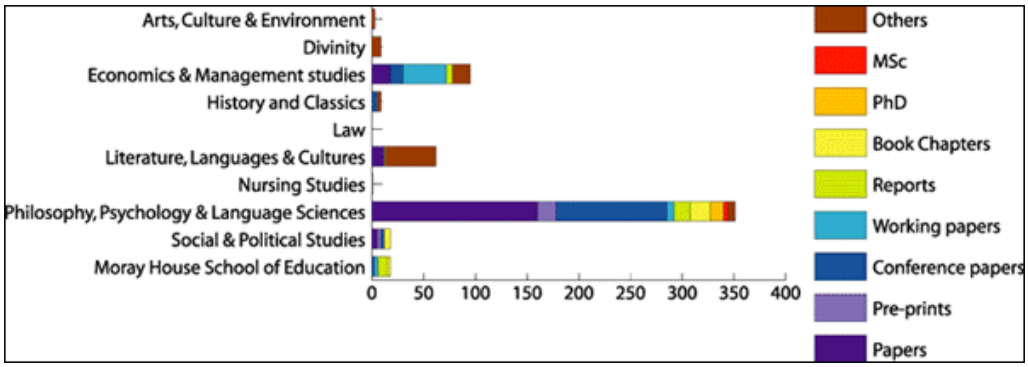

Figure 8: Volume and type of research material presently available in the HSS ed.ac.uk domain.

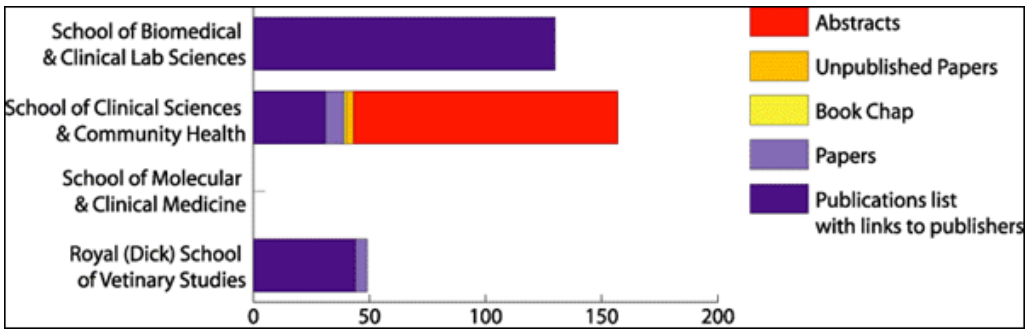

Figure 9: Volume and type of research material presently available in the MVM ed.ac.uk domain

8.5 Conclusions

The quantifiable data reviewed above supports the following broad conclusions:

- The number of e-print repositories has grown reasonably quickly but there has been a problem is persuading faculty to populate them once launched. This is supported by accounts of those starting repositories. One sub-population that is more biddable is that of postgraduate students and indeed many institutions have made it compulsory for theses and dissertations to be deposited electronically.
- Looking specifically at IRs, the majority are clearly in a very early stage of development but even most of the longer-established sites have a relatively small number of documents compared to the research outputs of their institutions. The exceptions are CERN (though is this really an IR? – it is surely really a subject-based repository) and the NASA technical report repositories.
- IRs to date have largely replicated the subject bias found in the older subject-based archives, i.e. content is largely maths, physics, computer science and economics. This is unsurprising, of course, since these are the areas where academics are known to be willing to self-archive, presumably because of the existing preprint cultures in these disciplines.
- Eprints are currently a small fraction (~22%) of the content on IRs and postprints currently appear to be a small fraction of these e-prints (though we have not established statistics on this split). So IRs are not at present threatening commercial publishers. However, this may be just a reflection of the early growth status of most IRs, or of a decision not to tackle controversial rights issues at first.

At present it is hard to tell whether IRs will follow the growth path of the subject-based repositories.

- There is little support in this evidence for IRs leading the reform or disaggregation of scholarly publishing.
- It is not clear what IRs are currently adding to the long-term preservation agenda, given their patchy coverage.

9 Publishers

In order to add to the list of issues for publishers described (by academics and librarians) in the literature reviewed, we decided to carry out a short online questionnaire of publishers and to hold informal discussions with some informed and involved publishers.

9.1 Questionnaire

A brief online questionnaire was designed and members of ALPSP and PA/CAPP were invited by email to complete it. (A copy of the questionnaire and the detailed responses are attached in the Appendices.) The total number of responses, 69 is a reasonably small proportion of the total combined membership (about 280), and the questionnaire was deliberately anonymous which could have allowed multiple responses from the same company (though multiple responses from the same computer were disallowed). Consequently the data gathered should be regarded as indicative rather than statistically significant, but may nonetheless be interesting.

A substantial majority of publishers expressing a view thought that IRs would have a significant impact on scholarly publishing, with 44% agreeing to 18% disagreeing, but 38% felt unable to answer either because they did not know enough about the issues or because the issue was currently too hard to judge.

Interestingly, a very large majority – 74% – thought that the impact on publishing would either be neutral (negatives balanced by positives) or there would be no significant impact. Only 19% thought the impact would be negative while 8% (i.e. five respondents) thought the impact would be positive.

Respondents were next invited to state unprompted what issues they thought were raised for publishers by IRs. The full list of responses is given in the appendix. The issues raised included rights, multiple versions, impact on the subscription model and the journals business models generally (e.g. loss of reprint income in clinical journals), diversion of library funds from acquisitions into building IRs, loss of article submissions to journals, etc. These points are included in the discussion of publishers' issues below.

Following completion of the unprompted question on issues raised for publishers, respondents were asked to rate a number of possible outcomes in terms of their likely impact, using a five-point scale. Publishers generally, but by no means unanimously, remain relaxed about the wide availability of *preprints* with 56% seeing this having a low or very low impact, and only 15% a high/very high impact. However they are (unsurprisingly) much more worried about the free availability of the final published and edited versions, with 41% believing this would have a high/very high impact compared to 21% for low/very low. A majority thought IRs could lead to a lowering of average quality owing to bypassing of publishers' peer-review and quality control. All respondents to this question thought that multiple versions would be a problem, with 78% believing this would have a high/very high impact. (This is of course a problem of pre- rather than post-prints.) Publishers appear undecided about how much IRs will weaken the roles of journals, and responses to the possible creation of new commercial roles for publishers were equally broadly spread.

Publishers were asked about their overall stance towards IRs. Responses were fairly equally split between a "wait and see" policy (40%) and one of active experimentation or collaboration (42%).

They were then asked (under the heading "Copyright issues") about their position on journal authors posting the final published version of their articles to IRs and e-print archives. 55% of the respondents currently permit and expect to continue to permit, while 24% currently prohibit and expect to continue to prohibit (reflecting the results of the RoMEO and ALPSP (Cox and Cox 2003) studies). Interestingly, some (12%) currently permit posting but expect

to restrict to exclude IRs or e-print archives in future, while 10% currently prohibit but expect to relax this prohibition in the future.

9.2 Issues raised for publishers by IRs

9.2.1 IPR/copyright

One of the more obvious issues for publishers is deciding how to respond to demands for more liberal copyright terms. As the RoMEO study showed (see above) there is by no means unanimity among publishers, particularly on the right to self-archive the final published version, as opposed to preprints. Some publishers make a distinction between the final peer-reviewed and edited version of the article and its typeset/laid-out version with the publisher's or journal's imprint.

There are however demands for even more liberal usage terms, again documented in the RoMEO study, for example for no restrictions (beyond observance of moral rights) on usage, including commercial re-use.

9.2.2 Appropriate copy

The appropriate copy problem for librarians has traditionally taken the form of how to prevent the library or its patrons repurchasing content (e.g. via document delivery services) that it has in fact already purchased (e.g. through a subscription or "big deal"). If the articles become freely available on IRs, then the "appropriate copy" for the library/user becomes the free one. Library software systems that can automatically redirect searches from publishers' paid-for content to a free equivalent thus undermine demand for the subscription.

9.2.3 Multiple versions

A related problem is the existence of multiple versions of an article. This is not specifically a problem for publishers alone of course: publishers and the academic community would generally agree that the definitive version was the peer-reviewed, edited version published in a journal, and most IR proponents would explicitly argue that this is not an aspect of the current scholarly publishing regime they wish to change. However for many purposes (i.e. the dissemination/communication part of journal publishing) the preprint is "good enough" and the widespread use of non-definitive (but free) versions may undermine the quality standards of journals and their importance.

9.2.4 Impact on journal subscriptions

There are a number of possible ways in which IRs might impact on journal subscriptions.

Free availability of the final article or near-final article might reduce demand for paid subscription access. This is of course the hope of the open access movement. The evidence to date (e.g. in physics) is that very comprehensive preprint archives (100% of the published literature is included in some areas of physics) can co-exist with subscription-model journals for over a decade. Advocates of open access (e.g. Crow 2002; Harnad 2001) argue that this is inconclusive and really just demonstrates the inertia in the current system, and indeed see the power of the institutions behind IRs as a key factor in breaking through this inertia.

Development of new open-access publishing on IRs might lead to the bypassing of (existing) journals altogether, i.e. particular journals might see their article submissions fall in favour of new non-subscription, IR-based competitors. This is really a variant on the launch of open-access journals, with IRs becoming publishers or possibly other entities acting as publishers by creating overlay journals on IRs. Dennis Hall, a UK physicist now associate provost at Vanderbilt University, argued in an October 2003 article (Hall 2003) that over the next five years the subscription model will be maintained but more and more universities will

experiment with IRs. Local quality control measures would then be likely to develop into broader peer review, thus "converting the institutional repository a de facto university-based journal".

Library funds might be diverted from content acquisition into building and maintaining IRs. Costs of building/maintaining IRs are discussed elsewhere in this report but it is interesting to note that the library at Georgia Tech was able to secure an increase in funding of almost $2m over two years to develop digital library services including an IR.

9.2.5 Future publishing models

It appears to be a stated aim of virtually all IR projects to reform the scholarly communication and scholarly publishing processes. Exactly what the replacement models might be is beyond the scope of this report but clearly it is an area for discussion.

9.2.6 Disaggregation of journal (publisher) functions

One of the platforms of the SPARC position is the idea that open access IRs will allow the journal publishing functions to be disaggregated and shared out among new players. Under this model, peer review, editing and formatting, dissemination, awareness, and archiving as well as the formal functions of registration (recording the author's precedence) and certification need not be carried by the same players as today. Some activists see peer review, perhaps with editing and formatting, as being the only functions that irreducibly belong to the publisher.

9.2.7 Integration of IRs and journals

Publishers are interested in exploring both how IRs can coexist with journals and – more creatively – how they can be integrated. Various possibilities exist, for instance:

- publishers could make their bibliographic data available through OAI-MHP. IOP Publishing has taken this step, which means that articles can be found transparently by anyone conducting a search on an appropriate OAI service provider (such as OAIster). Access to the full text is still, however, controlled under the terms of IOPP subscription arrangements;
- publishers could create overlay journals with IRs (and/or disciplinary repositories; there is little formal difference from an OAI perspective). Two existing examples are *Annals of Mathematics* (an overlay to arXiv) and *Perspectives in Electronic Publishing* (an experimental project at the University of Southampton);
- publishers can harvest OAI metadata (as can anyone else) as part of awareness or discovery services. One example is Elsevier's use of OAI data in its Scirus service;
- acceptance of IRs, and particularly their long-term preservation role, might accelerate the acceptance of all-electronic journals.

9.2.8 Opportunities for collaboration

What opportunities for collaboration currently exist between publishers and IRs? One example is at Oxford: the OU Library is setting up an IR (using EPrints software) as part of the SHERPA project. Oxford University Press will provide online access to articles by Oxford University-based authors published in many of the Oxford Journals from 2002, which will be searchable via the IR and freely accessible to researchers across the globe.

9.2.9 University Presses

With a few notable exceptions university presses are unprofitable and have to be subsidised by their host institutions. There are about 100 university presses in North America with a combined output of 10,000 books and 700 journals, and nearly all of these run at a loss. For a small press, the subsidy can be as large as 50% of the press's operating budget or of the order of $10,000 per book (Hall 2003). Given the demands on their finances, universities may balk at continued subsidy of their presses once (if) IRs become successful (well-used) but expensive. Raym Crow states in the SPARC position paper (Crow 2002) "This model recognizes that not all university presses would necessarily survive the proliferation of institutional repositories, as universities might logically consider the repositories a more efficient investment in scholarly communications than the universities' presses have traditionally been."

10 Conclusions

The case for the benefits to a research organisation of an institutional repository providing a set of infrastructural digital services including uploading/hosting, organising (metadata), disseminating and long-term preservation seems compelling. Most universities of any substantial scale do now appear to be either implementing or considering implementing such a repository, and funding bodies throughout the world are supporting research into their development.

What is far less clear is whether IRs will develop large, interoperable collections of *published* literature, as hope the advocates of open access. IRs are currently at an embryonic stage with only small, experimental collections of documents, but a clear message from the IRs is that one major hurdle – possibly the major hurdle – is overcoming faculty's inertia or indifference to self-archiving. It seems possible at present that IRs *per se* will fulfil a real and valuable function in supporting scholarly communication, research and teaching but that this function will be complementary to scholarly publishing rather in conflict with it. The impact of the wider open access movement is of course another matter.

Publishers in our survey appeared to share this view, as the large majority believed that IRs would have either no impact or a neutral one on scholarly publishing. Nonetheless there are clearly substantial challenges both in dealing with the wider issue of open access, of which IRs form a part, and in responding to the specific opportunities and issues raised by IRs.

11 Appendices

11.1 References

Andrew, Theo. (2003). Trends in Self-Posting of Research Material Online by Academic Staff. *Ariadne* **37**, October 2003. http://www.ariadne.ac.uk/issue37/andrew/

Barton, Mary R. and Walker, Judith H. (2002). MIT Libraries' DSpace Business Plan Project: Final Report to the Andrew W. Mellon Foundation July 2002. Available on the MIT DSpace site http://libraries.mit.edu/dspace-fed-test/implement/mellon.pdf

BOAI (2003). Budapest Open Archive Initiative Guide to Institutional Repository Software. First Edition, October 2003. http://www.soros.org/openaccess/software/

Branin, Joseph. (2002). A Proposal for Development of an OSU Knowledge Bank. Final report of the OSU Knowledge Bank Planning Committee. http://www.lib.ohio-state.edu/Lib_Info/scholarcom/KBproposal.html

CARL (2003). Canadian Association of Research Libraries: Institutional Repositories Position Statement. November 2003. CARL.

Cox, John and Cox, Laura. (2003). Scholarly Publishing Practice: the ALPSP report on academic journal publishers' policies and practices in online publishing. ALPSP. http://www.alpsp.org/publications/pub7.htm

Crow, Raym. (2002). The Case for Institutional Repositories: A SPARC position paper. Available from the SPARC website at http://www.arl.org/sparc/IR/ir.html

Day, Michael. (2003). Prospects for institutional e-print repositories in the United Kingdom (v1 28 May 2003). Available from the UKOLN website at http://www.ukoln.ac.uk/

ECS. (2003). Spreadsheet data on growth of EPrints and other repositories, available from the ECS, Southampton website at http://www.ecs.soton.ac.uk/~tdb01r/reports/2003-02-18-eprints_org_growth.xls

Gadd, E., Oppenheim, C. and Probets, S. (2003). The Intellectual Property Rights Issues Facing Self-archiving: Key Findings of the RoMEO Project. *D-Lib Magazine* **9(9)** September 2003. http://www.dlib.org/dlib/september03/gadd/09gadd.html. See also the six detailed RoMEO Studies 1-6 available from the RoMEO website and published in *Journal of Documentation*, *Journal of Information Science*, *Journal of Librarianship and Information Science*, *Learned Publishing*, and *Program*.

Hall, Dennis G. (2003). Some thoughts on journal publishing in the 21st century. Optics & Photonics News October 2003, pp30-33. Also available on the Kentucky, Tennessee, Knoxville and Vanderbilt Information Alliance website at http://www.lib.utk.edu/~alliance/Dennis_s_article.pdf

Harnad, Stevan. (2001). For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now. Available from CogPrints repository at http://cogprints.ecs.soton.ac.uk/archive/00001639/index.html

Hubbard, Bill. (2003). SHERPA and Institutional Repositories. *Serials* **16(3)** 2003, pp 243-247. Also on the Nottingham eprint repository at http://eprints.nottingham.ac.uk/archive/00000095/01/sherpa&instrep.pdf

Lynch, Clifford. (2003). Institutional Repositories: Essential infrastructure for scholarship in the digital age. ARL Bimonthly Report 226 (Feb 2003). Also available from the ARL website at http://www.arl.org/newsltr/226/ir.html

Martin, Ruth. (2003). ePrints UK: Developing a national e-prints archive. *Ariadne* **35** April 2003. http://www.ariadne.ac.uk/issue35/martin/intro.html

MIT (2003) MIT DSpace—A Case Study. http://dspace.org/implement/case-study.pdf

Nixon, William. (2003). DAEDALUS: Initial experiences with EPrints and DSpace at the University of Glasgow. *Ariadne* Issue **37**, October 2003. http://www.ariadne.ac.uk/issue37/nixon/intro.html

Pinfield, Stephen. (2003). Open Archives and UK Institutions. *D-Lib Magazine* **9(3)** March 2003. http://www.dlib.org/dlib/march03/pinfield/03pinfield.html

Smith, Abby. (2003). New-Model Scholarship: How Will It Survive? Council on Library and Information Resources, Washington, D.C. March 2003. http://www.clir.org/pubs/abstract/pub114abst.html

Van der Vaart, L. (2002). DARE: A New Age in the Provision of Academic Information. *D-Lib Magazine* **9(1)** January 2002. http://www.dlib.org/dlib/december02/12inbrief.html#VANDERVAART

11.2 Websites cited in the report

This table includes websites for the organisations and projects cited in the report except for the list of institutional repositories, which is given in the following section (11.3).

Organisation / project	*URL*
ARNO	http://www.uba.uva.nl/arno
bepress	http://www.bepress.com/repositories.html
Budapest OAI Guide to Institutional Repository Software	http://www.soros.org/openaccess/software/
CARL (Canadian Association of Research Libraries)	http://www.carl-abrc.ca
CDSware	http://cdsware.cern.ch/
CNRI Handle system	http://www.handle.net
Cogprints	http://cogprints.ecs.soton.ac.uk/
Creative Commons	http://creativecommons.org/
DAEDALUS	http://www.lib.gla.ac.uk/daedalus/
DSpace Federation, DSpace software	http://www.dspace.org
Ebrary	http://www.ebrary.com/libraries/ir.jsp
ECS - Electronic and Computer Science eprints	http://eprints.ecs.soton.ac.uk/
ePrints UK Project	http://www.rdn.ac.uk/projects/eprints-uk/
Eprints.org, Eprints software	http://www.eprints.org
Fedora	http://www.fedora.info/
Ingenta	http://www.ingenta.com
i-Tor	http://www.i-tor.org/en/toon
JISC Focus on Access to Institutional Resources	http://www.jisc.ac.uk/index.cfm?name=programme_fair
Kepler	http://dlib.cs.odu.edu/
Lund eprint server	http://eprints.lub.lu.se/
MIT OpenCourseWare	http://ocw.mit.edu/index.html
MPG eDoc	http://edoc.mpg.de/doc/help/aboutus.epl
MyCoRe	http://www.mycore.de/engl/index.html
Open Archival Information System Reference Model	http://ssdoo.gsfc.nasa.gov/nost/isoas/
Open Archive Initiative	http://www.openarchives.org
Open Digital Rights Language Initiative	http://odrl.net/
OpenURL Framework	http://www.niso.org/committees/committee_ax.html
OPUS	http://elib.uni-stuttgart.de/opus/doku/english/about_english.php
PubMedCentral	http://www.pubmedcentral.nih.gov/
RoMEO: Rights Metadata for Open Archiving	http://www.jisc.ac.uk/index.cfm?name=project_fair_romeo
SHERPA	http://www.sherpa.ac.uk/
SPARC Europe - Institutional Repositories	http://www.sparceurope.org/Repositories/
SPARC US - Institutional Repositories	http://www.arl.org/sparc/core/index.asp?page=m0
TARDis	http://tardis.eprints.org/

11.3 Table of Institutional Repositories

Country	*Site*	*URL*
Australia	Australia National University Eprints Repository	http://eprints.anu.edu.au/
Australia	Queensland University eprints@UQ	http://eprint.uq.edu.au
Austria	Sammelpunkt. Elektronisch archivierte Theorie (Univ Vienna)	http://sammelpunkt.philo.at
Belgium	Louvain Catholic Univ	http://edoc.bib.ucl.ac.be/index_ucl_epr.html
Canada	Université de Montréal Papyrus - Institutional Eprints Repository	http://papyrus.bib.umontreal.ca/
Denmark	Aalborg University Electronic Library	http://www.aub.auc.dk/phd/mainpage.html
France	@rchiveSIC Archive Ouverte en Sciences de l'Information et de la Communication	http://archivesic.ccsd.cnrs.fr/
France	ENS LSH (Econle Normale Superieur, Lettres et Science Humaines)	http://eprints.ens-lsh.fr/
France	Institut Jean Nicod Archive Electronique	http://jeannicod.ccsd.cnrs.fr/
Germany	Dresden - HSSS	http://hsss.slub-dresden.de/hsss/oai/
Germany	Universität Dortmund: Eldorado	http://eldorado.uni-dortmund.de:8080/
Germany	Universität Essen: MILESS	http://miless.uni-essen.de/
Germany	Universität Konstanz: KOPS-Datenbank Konstanzer Online-Publikations-System	http://www.ub.uni-konstanz.de/kops/
Germany	Universität Stuttgart: OPUS (Online Publications University of Stuttgart)	http://elib.uni-stuttgart.de/opus/doku/english/index.html
India	Indian Institute of Science (Bangalore)	http://eprints.iisc.ernet.in/
Ireland	National University of Ireland, Maynooth	http://eprints.may.ie/
Israel	WISDOM - Weizmann Institute, Israel	http://wisdomarchive.wisdom.weizmann.ac.il:81/
Italy	UNITN-eprints : University of Trento - Italy	http://eprints.biblio.unitn.it/
Italy	Università degli studi di Firenze	http://e-prints.unifi.it/
Italy	University of Firenze eprints	http://biblio.unifi.it
Netherlands	University of Maastricht	http://137.120.22.236/www-edocs/default.asp?taal=ENG&webnaam=edocs
Netherlands	Utrecht University: Dispute	http://dispute.library.uu.nl/
Slovenia	University of Ljubljana, Slovenia.	http://eprints.fri.uni-lj.si
Sweden	Blekinge Institute of Technology: Electronic Research Archive	http://www.hk-r.se/fou/
Sweden	Lulea Institute of Technology	http://epubl.luth.se/index-en.html
Sweden	Lunds Universitet	http://www.lub.lu.se/luft/
Sweden	Uppsala University	http://publications.uu.se/index.xsql?lang=en
Switzerland	CERN Scientific Information Service	http://cds.cern.ch/
UK	Dspace @ Cambridge	http://www.lib.cam.ac.uk/dspace/
UK	University of Bath: ePrints@Bath	http://eprints.bath.ac.uk/
UK	University of Glasgow: EPrints at Glasgow	http://eprints.lib.gla.ac.uk/
UK	University of Nottingham	http://www-db.library.nottingham.ac.uk/eprints/
UK	University of Strathclyde	http://eprints.cdlr.strath.ac.uk/
USA	California Digital Library: eScholarship	http://escholarship.cdlib.org/wprepositories.html
USA	Caltech CODA: Caltech Collection of Open Digital Archives	http://coda.caltech.edu
USA	D-Scholarship, Florida State Univ	http://dscholarship.lib.fsu.edu/
USA	Georgia Tech IR	n/a
USA	Hofstra University: HofPrints	http://hofprints.hofstra.edu/
USA	MIT Dspace	https://hpds1.mit.edu/index.jsp
USA	NASA - GENESIS EPrints	http://genesis2.jpl.nasa.gov/
USA	NASA - Research Institute for Advanced Computer Science	http://eprints.riacs.edu
USA	Ohio State Univ KnowledgeBank	https://dspace.lib.ohio-state.edu/index.jsp http://www.lib.ohio-state.edu/KBinfo/
USA	U Virginia (Fedora)	n/a
USA	Virginia Polytechnic, Digital Library and Archives	http://scholar.lib.vt.edu/DLASPS/index.html
USA	Virginia Tech, VT CS Technical Reports	http://eprints.cs.vt.edu:8000/

11.4 Publisher questionnaire

1. Do you believe that the development of Institutional Repositories will have a significant impact on overall scholarly publishing within the next 5 years?						
	Response					
Yes	30	44%				
No	12	18%				
Don't know – too hard to judge	16	24%				
Don't know – I'm not familiar enough with the issues	10	15%				
Total Respondents	68					
(skipped this question)	1					
2. Do you think the commercial impact on your publishing business will on balance be						
	Response					
Positive	5	8%				
Negative	12	19%				
Neutral (i.e. balanced positive and negative impacts)	37	58%				
No significant impact	10	16%				
Total Respondents	64					
(skipped this question)	5					
3. What issues for publishers do you think are raised by IRs?						
Total Respondents	41		See below			
(skipped this question)	28					

4. Please rate the following possible outcomes according to their likely impact						
	Very high	*High*	*Medium*	*Low*	*Very low*	*Response Average*
Cancellation of subscriptions owing to free availability of preprints in IRs	1	7	13	19	7	3.51
	2%	*15%*	*28%*	*40%*	*15%*	
Cancellation of subscriptions owing to free availability of final published and edited articles in IRs	6	14	17	9	2	2.73
	13%	*29%*	*35%*	*19%*	*4%*	
Lowering of average quality owing to bypassing of publishers' peer-review and quality control	11	12	16	8	1	2.5
	23%	*25%*	*33%*	*17%*	*2%*	
Proliferation of multiple versions of articles	15	22	10	1	0	1.94
	31%	*46%*	*21%*	*2%*	*0%*	
Weakening of the role played by journals in scholarly publishing	4	11	15	16	2	3.02
	8%	*23%*	*31%*	*33%*	*4%*	
Creation of new commercial roles for publishers	3	9	25	8	2	2.94
	6%	*19%*	*53%*	*17%*	*4%*	
Total Respondents	48					
(skipped this question)	21					
5. What is your company's overall stance towards IRs?						
	Response					
Wait and see	19	40%				
Active experimentation or collaboration	20	42%				
Neutral – no real relevance to our business	4	8%				

Not yet got round to developing a stance	5	10%				
Total Respondents	48					
(skipped this question)	21					
6. What is your position on journal authors posting the final published version of their articles to IRs and eprint archives?						
	Response					
Currently permit posting and expect to continue to do so	23	55%				
Currently permit posting but expect to restrict to exclude IRs or eprint archives in future	5	12%				
Currently prohibit posting to IRs/archives and expect to continue this prohibition	10	24%				
Currently prohibit but expect to relax this restriction in the future	4	10%				
Total Respondents	42					
(skipped this question)	25					
7. Please rate the following topics in terms of their interest to you as potential conference presentations						
	1 (=Very interested)	*2*	*3*	*4*	*5 (Not at all interested)*	*Response Average*
Case studies from IR project managers (e.g. DSpace Eprints etc.)	15	13	12	7	0	2.23
	32%	*28%*	*26%*	*15%*	*0%*	
Presentation on functionality and technical capabilities of current IR software	3	12	13	13	5	3.11
	7%	*26%*	*28%*	*28%*	*11%*	
Copyright and IPR issues raised by IRs	14	16	10	4	3	2.28
	30%	*34%*	*21%*	*9%*	*6%*	

Reports from JISC FAIR (Focus on Access to Institutional Resources) projects	7	10	17	7	4	2.8
	16%	*22%*	*38%*	*16%*	*9%*	
Implementation issues faced by those setting up IRs	8	11	12	7	8	2.91
	17%	*24%*	*26%*	*15%*	*17%*	
Costing / funding models for IRs	17	17	9	3	1	2.02
	36%	*36%*	*19%*	*6%*	*2%*	
High-level review of potential impact on scholarly communication given by publisher	29	16	1	0	0	1.39
	63%	*35%*	*2%*	*0%*	*0%*	
High-level review of potential impact on scholarly communication given by advocate of new publishing models	22	19	5	1	0	1.68
	47%	*40%*	*11%*	*2%*	*0%*	
Case studies by publishers working with IRs or eprint archives	20	22	4	1	0	1.7
	43%	*47%*	*9%*	*2%*	*0%*	
Presentation of research into how authors are responding to IRs	21	18	7	1	0	1.74
	45%	*38%*	*15%*	*2%*	*0%*	
Total Respondents	47					
(skipped this question)	22					

3. What issues for publishers do you think are raised by IRs?	
1.	As a non-profit university press, question #6 is difficult to answer as our relationship with our journals is quite different from that of a commercial press. We do not own copyright on any of our journals - it is held either by the journal itself or an association that may be behind it. Therefore decisions regarding publishing a final article in an institutional repository must be made by those groups. Our press could take a stand on encouraging such a position but has not done so yet. The e-journals we currently publish are already open access so there should be no problem with articles also being in the repository. We are currently exploring having a brand-new journal actually use the repository infrastructure under our Press imprint
2.	Adjustment of copyright transfer and regulations Reduction of submitted papers Reduction of subscription income P.S. We are not commercial publishers ourselves
3.	As a book publisher my knowledge is limited but for what it's worth: 1 Open access is something that's been talked about for a quite a few years and so far there hasn't been a significant impact on commercial publishers, which indicates that it's not so easy for institutions to implement 2 this provides opportunities for publishers to collaborate on open access/author pays models; BiomedCentral have started but not the big (Index Medicus cited) journals. 3 is there any evidence of author perception of publication status in IRs? 4 is the current estimated cost of author pays (£1500 v £4500 for buying the journal - ref current BMJ editorial) likely remain cheaper than the traditional method? 5 will IRs serve the whole community effectively? Is this something publishers can help institutions with? 6 reuse of information: can publishers help institutions with this?
4.	One main issue: how to develop value-added elements so as to keep in business. The role of publishers is no longer selfevident. So, perhaps: - rigorous peer review & selection, so that researchers save time by relying on publishers as intermediates - high-quality production and editing with good add-ons, so as to make materials most accessible and pleasant to work with - providing context, e.g. by adding backfiles Or seek some kind of cooperation with research institutions.
5.	Publishers may be cut out of the peer review process. Scholars in our field worry about the reliability of scholarly communications without peer review and editorial controls. Another concern is copyright. Some academic institutions want to own the work of their faculty members, making it unlikely that publishers could acquire and publish the work without cost.
6.	Most important, the peer-reviewed version that a publisher issues must be differentiated from other versions at IRs or at the author's website.
7.	Keeping track with institutional requirements, the developing usage by library patrons, security issues, guarding against obsolescence, ensuring that any collaboration is mutually beneficial and meets scholarly requirements of libraries and business needs of publishers.
8.	For journals publishers, there is the question of whether IRs will provide a sufficiently attractive alternative to make it more difficult to acquire high-quality articles. I doubt that this will happen in the near future, but it is a possibility. For book publishers, the issues are not so clear. So far, IRs are publishing mostly article-length material, but there is the possibility that they could also attract book-length material. This would probably not have a negative impact on publishers, because the material published in IRs is likely to be that which is not commercially viable, even for university and other nonprofit presses.
9.	Publishers must emphasize that the advantages of scholarly publication through traditional methods are peer review, close editing, and high-quality presentation. Publishers should point out the distinction between the kind of material published by IRs and that published by traditional scholarly publishers.
10.	Potential loss of copyright control, loss of sublicensing and permissions revenue, loss of sales for backlist steady-sellers, reduction in storage costs for archives and old inventory, reduction in costs of keeping books in print (if IRs provide print-on-demand capability).
11.	Controls on linking/searching/spidering across institutions.

12.	1. Competitive pressure from the organization of peer-reviewed articles in open-access, interlinked collections, with full-text indexing for popular free web engine free and retrieval. 2. Re-raises the liberal principles of author retained posting rights widely adopted in many copyright policies in recent years. 3. May put more pressure on Digital Library offerings than on individual journal titles, oddly pushing publishers back towards print only, since the imprimatur is still valued and not supplied by IRs.
13.	Potential loss of control over access to the authoritative version of a work. Potential harm to publishers if all or most of their publications come from institutions with repositories (this situation is NOT the case in the fields where we publish).
14.	1 That the cost to the institution of open access is not fully audited and transparently presented so that appropriate comparisons can be made with the cost of acquiring paid-for content. Particularly important as publishers continue to develop business models which offer greater flexibility and value for money for institutions in terms of purchasing and content options. 2 Users will be able to search across meta level databases that include details about content that is both free and toll access. Metadata for different versions of the same article will sit side by side. Users may not appreciate the value-added version available via the publisher and 'replace' it with the free access version. Over time that undermines the publishers' business model - lower revenues may mean insufficient cash to cover the costs of publication, leading to the potential demise of well established, previously valued (by the faculty) journals. For many learned societies the surplus generated by their journal(s) is used to fund services for members, research, facilities, etc. 3 Not all subject areas are in a position to fund open access from research/institutional funds - eg the LIS and most social science areas. What will be the effects in this area? 4 Not all regions of the world will be able to fund the development of IRs - this potentially disenfranchises the lesser developed countries - who are benefitting from many arrangements by publishers to gain free/virtually free access to published content. 5 A downgrading of the quality of scholarly content as more and more becomes available via IRs - who by definition will 'publish' everything produced by their institutions. We already know that users have little ability to discern quality of content available via the web (a major concern of librarians). 6 Technically it can be done (although in practice advocacy programmes are necessary as authors do not necessarily see the benefit) but that doesn't mean to say it's a good idea! Supplier driven initiatives are riddled with inefficiency. Intiatives driven by the market tend to become more efficient over time and sustained. 7 The model presented by SPARC for the transition of existing journals from sub-based to open access is unrealistic. I've tried it (with cooperation from SPARC and Sconul) and the librarians weren't interested. Where do we go from here? 8 We are not convinced that it's what authors want in reality. Nor that the potential implications are fully understood by the various stakeholders involved in scholarly communication. A great deal of assumption and naivety has been seen. Agree that sub costs are continuing to rise and in the face of that something needs to be done. But in effect it is already happening as publishers present different and evolving access/purchase models. 9 Lack of clarity in terms of which is the definitive version of the article. 10 Lack of appropriate peer review. 11 The Zwolle Principles regarding copyright (to which we were the first publisher to sign up to) support the principle that it is important for each stakeholder to have the appropriate rights, rather than for a particular stakeholder to hold the copyright. An important point - eg publishers are in a better position to achieve the authors' goals if they hold the copyright rather than the author. The author needs rights, not necessarily the copyright.
15.	1. Resource costs 2. Development management 3. Revenue generation
16.	coordinated IRs could be dangerous for some journals, i.e. in physics where there is already a strong tradition of cooperation in sharing information amongst academic researchers;
17.	Version control - less need for updated editions etc. Demise of some smaller subscription based learned societies/publishers Joint projects with IRs (OUP-Bodley's SHERPA) - possibly provides publisher with resource discovery tool for additions to list.
18.	Economic issues mainly - the open access business model is far from proven. Copyright will also be an issue.
19.	May perhaps increase the pressure on publishers to move to Open Access model for journals, as journals become forced to compete with free repositories for authors. Authors (and their funding bodies) may well still be willing to pay for OA journal publication because of added value offered: most importantly prestige, but also

	functionality, linkage (especially via reference lists), formatting, indexing. (Peer review is still important, at least in life sciences, but mostly benefits readers, although also benefits authors indirectly in maintaining prestige.) Authors will continue to submit to subscription-only journals (as can be seen from high-energy physics) but how long will libaries continue to support subscriptions, when readers can note who has been published there and then search for the same article via interoperable repositories for free? Perhaps for a while, as for high-energy physics, but surely not indefinitely...
20.	Who owns the DOI? Can repositories expect to post the publisher's final version of a paper or only a pre-print, i.e. before the publisher has added value?
21.	IT WILL BE USEFUL POOL FOR ALL USERS OF THE JOURNALS CONTAINED IN THERE
22.	A&I General author attitudes Nothing specific that isn't raised by other initiatives (e.g. open access)
23.	1.Rethinking their business model made more urgent- need to find middle way or viable transition 2.as ever, need to promote thinking from more than just library or author approach
24.	GOOD SOURCE AS THE LOCAL LEGAL DEPOSIT
25.	current subscription model in jeopardy if researchers can access research articles for free using good search engines to reach institutional repositories what rights authors retain
26.	Control of IP Author relations Institutional relations
27.	The worry is that people will move away from publishing in our journals and opt for IRs. I have doubts that this will happen short term, as our journals have a stamp of quality (our refereeing) procedure and editorial quality, but I suspect we'll have to keep on our toes to keep this. I don't really know enough about this, I suspect it'll first have impacts in the sciences and only slowly move to humanities (if at all).
28.	There will be reduced submissions to journals if Institutions insist that their staff use the repositories. Although we do allow authors to post their articles on their own website we are not in favour of publishing articles that are posted in a networked environment.
29.	format and methods of capturing content
30.	Control over the rights to an article; clarity as to the definitive original copy of an article; legal uncertainties if there is something wrong with the article; loss of reprint revenue which will be significant for clinical journals.
31.	business models access rights
32.	Relationship between the role of peer-reviewed journals and IRs Can publishers play any role in the development and managemnet of IRs
33.	1. Diversion of money that might otherwise have been spent on purchasing journals, books and other online resources
34.	Validated and approved final copy confusion. Also, of course the impact on traditional business models.
35.	If the papers are duplicates of published papers, users could be attracted to these copies rather than the journal's site. This would affect subs, the journal's brand and visibility, potential for online advertising (fewer views), and so on. If they are not, some form of peer review will probably still be desired. Publishers may find themselves in competition with new bodies offering peer review only, which they could do without incurring the costs of distribution. In an Open Access world, such bodies would be able to compete on price very effectively.
36.	Copyright Assignments and the rights granted back to the author; The definition of "prior publication"; As a secondary publisher, the need (or otherwise) to include deposited content in our database; The relationship between the IR and the published journal - will it be a comfortable one or not?
37.	Potential loss of subscription income. Potential loss of authors. Intellectual property rights. Possible erosion of the strength of journal 'brands'.

38.	Confusion regarding the authentic copy. Intellectual Property Rights - may play a part in universities trying to get mor control of these from authors. Confusion re branding. Money diverted to universities trying to do all this content management instead of spending it on our stuff.
39.	ownership of material and the ability to develop it. Scope for misunderstanding between publishers and institutional employees with regard for new projects. Need for industry clarity and consensus, and necessity to communicate this to institutions clearly
40.	Copyright Appropriate copy of the article

11.5 eScholarship Policy

Taken from the eScholarship website 8 January 2003

11.5.1 Who Can Join

Any University of California research unit (ORU or MRU), institute, center, or department is elegible to join. A UC unit is one governed by the University of California Regents.

11.5.2 Whose Papers Can Be Included in the Repository

Content does not have to be authored by UC faculty to be included in the eScholarship Repository. For example, a unit may use the repository to post papers from a conference they sponsor, which includes faculty from UC and other institutions. All that is required is that the sponsoring unit decides that the content is appropriate for the repository.

11.5.3 Appropriate Submissions

Any content is appropriate if all applicable policies are followed (e.g., copyright), it is technically feasible (the content can be posted using existing format types, etc.), and the sponsoring unit decides it is appropriate. We do not accommodate the posting of bibliographic citations or abstracts alone, without the referenced paper. If you have any questions, please contact us at help@repositories.cdlib.org.

11.5.4 Peer-Reviewed Series

The eScholarship Repository infrastructure also supports peer-reviewed series and journals. If you are interested in using the repository for peer-reviewed content, visit our information page, which will help you decide whether this is the right forum for your scholarship. Your campus eScholarship liaison is also a useful resource.

11.5.5 Removing a Paper

Authors may request that the unit system administrator remove their paper, or a version of their paper. However, once a paper is deposited in the repository, a citation to the paper will always remain. The exception is peer-reviewed series and journals, where removal is not allowed.

For example, if an author decides they don't want a working paper to appear on the repository anymore, they ask the system administrator at their unit to remove the paper, which hides it from public view. Instead of the paper appearing in the repository, there is instead a citation saying that this paper - by this person, published on this date, with this URL - has been removed. This means the URL never disappears, though a paper may be removed.

The repository allows faculty to show the progression of their research, should they so desire. Ten different versions of papers could be posted on the repository, with all of them visible. Or the faculty member could ask the repository administrator to remove the 9 earlier versions, leaving only the most recent one visible. However, in addition to the current version, there would be 9 citations showing that there had been 9 earlier versions available, published on these dates, with these titles, etc.

If a paper is being removed because of subsequent journal publication, please consult the Copyright section below.

11.5.6 Author Review

This is a step whereby authors are given the opportunity to review the PDF after the paper has been uploaded to the system but before it is posted. Since the system can automatically create a PDF from a Word or RTF document, in some cases it's especially important that the author check the PDF one more time. It is up to each unit whether or not they want to have author review. The exception is peer-reviewed series and journals, where author review is required.

11.5.7 Author Agreements

In the agreement signed by the unit director or department chair, the participating unit guarantees that they will obtain certain assurances from their authors. Suggested language for an author agreement is provided.

11.5.8 Copyright

Authors retain the copyright for all content posted in the repository. The author agreement specifies a nonexclusive right to use. This means the author is free to reuse the content elsewhere.

If a working paper is published in a journal—either in the same form or, more commonly, in revised form—many journals allow the working paper to continue to be made available, especially when it is for educational/scholarly noncommercial use. Unfortunately, some journals do require that the working paper be removed. Others grant exceptions for something like the eScholarship Repository; they just need to be asked. It is up to the faculty member to check the terms of their agreement with the journal to see what is allowed. Individual journal policies vary widely. The RoMEO Project (Rights MEtadata for Open archiving) has compiled a list of many journals' "Copyright Policies" about "self-archiving."

If you are interested in including a reprint of a journal article on your repository site, the faculty member should check their agreement with the journal to see if it is allowed. If it would not violate copyright, you're welcome to do so.

You are the gatekeeper for your repository site, and it is up to you to decide what is appropriate—as long as it doesn't violate copyright and conforms to eScholarship Repository policies.

For more information on copyright issues as they relate to the topic of reshaping scholarly communication, please see the UC Libraries site.