Data is an Asset

The Challenges and Benefits of Better Enterprise Data Management

Table of Contents

Data and Process Advantage Limited
Advantage House, Stowe Court, Stowe Street
Lichfield
Staffordshire
WS13 6AQ
England

Telephone: (+44) 1543 401144
Web site: www.dpadvantage.co.uk
Contact: enquiries@dpadvantage.co.uk

1 Introduction

This White Paper describes why data is an asset to an organisation and the benefits of treating enterprise data as an asset. We look at the challenges of managing enterprise data and explain why data quality management is such an important process in all large organisations, with illustrative examples of what can go wrong if data quality is not managed effectively. The paper finishes by looking at a framework for how to mitigate the risks of poor data which is explained in more detail in a separate White Paper.

2 Data is an asset

Data is being created at a faster rate than ever before - however you forecast future data growth, data is now more important than it ever has been. As the world becomes a more data driven place, smart businesses can gain competitive advantages by exploiting data more effectively. This vast data explosion brings newer, different challenges to businesses - It is one thing to store a lot of data, but the benefits will only come if the data is of suitable quality and reaches the right people at the right time. A mindset of treating data as an asset will help your organisation to achieve this.

Many larger organisations, such as utilities and transport, are developing management systems that provide more effective and sustainable management of their activities. Managing data requires a similar mindset.

An asset is a resource with value that can deliver benefit to an organisation, therefore data warrants being treated in the same way as a physical asset. Like physical assets, data:

- Can have high value for your organisation;
- Can drive up business performance and safety by enabling better informed decisions;
- May have legal or regulatory requirements to be managed effectively;
- Has a life cycle, from conception to capture to operation and renewal; and
- Can increase business costs if not managed effectively (and therefore reduce efficiency and profitability).

Unlike physical assets, data supports strategic decision making – get this wrong and you may end up making incorrect, potentially expensive, decisions that could have long term impacts for the organisation. Also, unlike physical assets, when the data asset is used, it is not consumed or destroyed – in fact, the more data is used, then arguably the more value it could generate.

Whatever sector your organisation operates in, there are benefits to be gained from treating your data as an asset to your organisation. This means thinking about the data that drives organisational decisions and activities, not just the software and applications that use it.

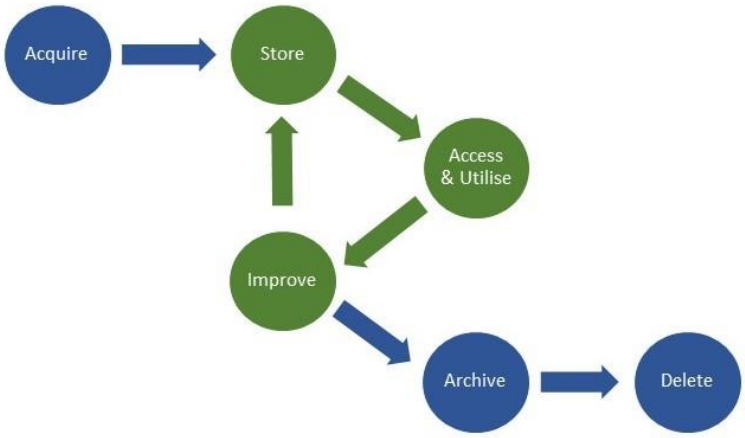
2.1 What type of asset is enterprise data?

The phrase "treat data like an asset" is used increasingly frequently, however, assets come in many forms. So, what type of asset is enterprise data?

Some assets may be large, robust, fixed assets that, once built, will exist for centuries, such as the Hoover Dam. However, data is not robust like this and perhaps needs to be considered more like a sandcastle – the individual grains of sand represent items of data, and the configuration of sand that makes the sandcastle is the information that has value to the organisation. A sandcastle is a fragile asset that can easily be degraded by wind and waves. Like a sandcastle, the data and information is a fragile asset that can easily be degraded by people, systems, and processes. Data risks losing its credibility if its condition is not monitored and the quality understood and nurtured. This may seem difficult to achieve, but that does not have to be the case.

2.2 The Data Life Cycle

Data, like other assets in your organisation, has a lifecycle. Therefore, the benefits of good quality data may be delivered in cycles or distinct phases from acquiring data, right through to eventual archive and deletion.



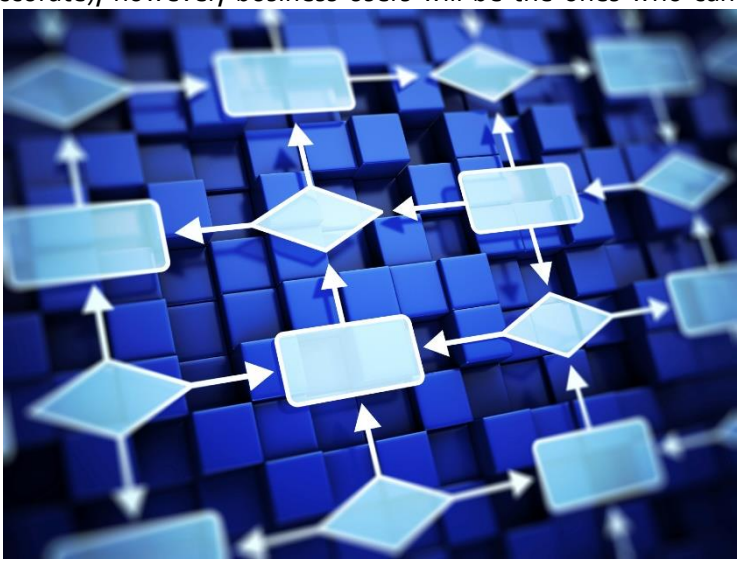
Acquire	•Specify what data you require •Collect or purchase this data
Store	•Store in agreed database
Access & utilise	•An input and output from processes •Enables use in decision support tools
Improve	•Correct identified data quality issues •Gather additional attributes when data requirements change
Archive	•Identify historic data that is not required in 'live' systems •Move data to off-line data storage
Delete	•Permanent deletion of data that is no longer of value to the organisation and has no legal requirement for retention

Throughout the data lifecycle, data can degrade in quality or become completely redundant if not monitored and maintained effectively. Such a long term, insidious decline in data quality can be very difficult to reverse – data backups will only take you to a version of the data from a day, week or month ago. However, if data quality has degraded over many months and years, then it will be almost impossible to reverse. This could pose a high cost & safety risk to your organisation, especially if you rely on data to make decisions, particularly in an environment where risk to personal health and safety is a significant factor.

3 The challenges of managing enterprise data

Larger organisations tend to be complex with many overlapping processes, functions, systems, and data sets. A single data field may be an input to (and output from) multiple processes, therefore, managing the quality of this data is more challenging since more factors may need to be considered. This is often even more relevant in organisations with a lot of legacy systems and data due to the nature of how systems may have developed in the past. 'Unofficial' systems, perhaps developed by users with spreadsheet based tools, further complicate this situation.

Often, people assume that 'IT' will look after the data, however, they will only generally be concerned about data that is invalid (incorrect formats) or where mandatory data is missing and therefore causing application errors. Data that is valid, but inaccurate, will generally not be a 'problem' for IT (who may possibly struggle to identify that it is inaccurate), however, business users will be the ones who can identify (and possibly correct) inaccurate data.



Some people suggest that you assign 'Data Owners', however, as has been observed, due to the many processes that may affect the quality of data, then the Data Owner may be in a relatively powerless position. Perhaps a better approach is to ensure that Process Owners are identified and that they help ensure that the data created/updated by the processes that they own is of suitable quality. This will then mean that if it is decided to assign Data Owners, that they have more control over the data for which they are responsible.

In organisations with several legacy systems and data stores, it can be more difficult to ensure you get the right data to the right people at the right time, to a high enough level of quality. The most common challenges to managing enterprise data are described in the following tables, and we have grouped these challenges into People, Systems & Processes challenges.

3.1 People challenges:

It is easy to think of data as the type of asset that is not affected by people, but of course human nature will always factor into any business process. Below are examples of how people and behaviour impact enterprise data:

People Challenges	Description	Issue
Diligence	People may have competing targets and pressures or may not appreciate the importance of data	Time and effort may not be spent on the data which may end up missing or poor
Competence	Lack of appropriate skills, knowledge, and training to support data management tasks	Experience and competencies of managing will vary across an organisation. Ensure people have appropriate skills and training to support data management tasks
Enthusiastic innovators	People with a tendency to always try and 'improve' how things are done	Unauthorised, but well intentioned, variations in approaches to approved methods
Local data stores	Staff with unofficial, local stores of information	Data in the corporate system may not reflect the true picture. Other staff may not be aware of these separate data sets
Different views	Some data attributes may be subjective with different views of what is correct	Poor data definitions and poor process may lead to 'flip-flop' data attributes

3.2 Technology challenges

Challenges: Business systems and technology will inevitably present challenges and opportunities to your data quality:

Technology Challenges	Description	Issue
Multiple systems	The same data entry may exist in more than one system	With no effective Master Data Management (MDM) it is harder to determine the correct data value
Complex data architecture	Data architecture dictated by the software tool	Vendors own data models have to be used which may make Master Data Management more challenging
Data silos	Data that cannot be stored in corporate system must be in separate data stores	Difficulty bring all data together. More challenges managing across data stores when data updates are required
Presentation masking poor data	New software may be more attractive than the previous version, but the data may still be poor	Users either do not spot the poor data. or criticise the software implementation for not fixing data errors known by the user

3.3 Process challenges

Processes, like the data itself, will need constant monitoring and nurturing so they remain fit for purpose, helping your data to do the same:

Process Challenges	Description	Issue
Overlapping processes	Different processes may use and update the same data entries	Difficulty identifying the cause of data errors. A challenge for data owners
Process understanding	Staff learning from colleagues due to lack of refresher and new-starter training	Lack of awareness of what data needs to be recorded and why. Incorrect process outcomes
Competing targets	Productivity targets may not allow time for complete data updates	Short term productivity gains at the expense of long term data degradation
Governance	Lack of effective control over process performance and change control	Data errors not spotted early. Inefficiencies and opportunities not addressed

The above could perhaps be summarised by the statement "*valid but inaccurate data can lead to false perceptions of reality.*"

4 What is Data Quality?

There are many definitions for data quality; one definition is:

The right data, available at the right time to the right users, to make the right decision and achieve the right outcome.

Whilst 'perfect' data quality may be desirable, the reality is that organisations will not have the time, resources, or budget to have 'perfect' data. Therefore, if organisations accept that their data quality is not perfect, they can then seek to understand and describe the nature of their data quality.

For example, if someone states that 'the weather is bad' this has little meaning without stating whether it is too hot/ too cold, too wet/ too dry, too windy/ too still etc. Similarly, if someone states that they have poor quality data, this can be difficult to interpret without a better way of describing the nature of your data quality. Therefore, it is useful to use different data quality attributes (or dimensions).

4.1 Data Quality Attributes

There are many different documented definitions for what constitutes the attributes of data quality, but the attributes that we find most useful are:

- Accuracy
- Completeness
- Consistency
- Validity
- Timeliness
- Uniqueness

Information in the following table of children's toy bricks helps illustrate these dimensions:

ID	Type	Length	Width	Height	Colour	Studs	Purchase date	Cost
010	Wood	59.5	29.0	29.0	Yellow	-		
012	Wood	59.5	28.9	28.9		N/A	01-09-2001	£8.42
014	Lego	79.8	31.8	9.6	Black	10x4		
015	Lego	31.8	15.8	11.4	Blue	4x2	12-23-91	£2
044	Lego	47.8	7.8	9.6	Grey	6x1	27/4/14	£7.12
045	Wood	60.0	29.5	28.6	Yellow		15/7/15	£4.21

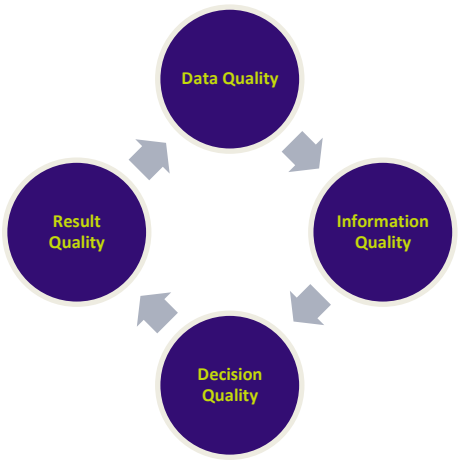
- **Accuracy -** how closely a record reflects the real object it represents. For example, looking at the data table above, by inspecting the real object (the bricks) we can verify whether Brick 045 is a yellow wooden block with the dimensions L 60xW 29.5xH 28.6 . If the real object turns out to be a green brick, or with different dimensions to those in the table, then the data is inaccurate;
- **Completeness** - A measure of whether all items recorded and all their attributes are populated. For example, the attributes for Brick 010 are not complete. Similarly, if the toy box contains a Brick 017, the list of bricks is not complete;

- **Consistency -** An entity recorded in more than one data store is comparable across data stores. For example, Block 015 from the table above has a purchase date of 01/09/2001, but in the purchasing system the transaction date may be 04/12/2001.
- **Validity -** Data conforms to the specified format. For example, the Purchase Date field contains many different date formats - which is the valid format?
- **Timeliness -** Data is up to date and is available to users in a timely manner. For example, the entry for Brick 045 may have been added 2 months after the purchase date, which may be slower than the required update frequency.
- **Uniqueness -** A single representation exists for each physical entity. For example, in the table above, no entry appears twice, therefore the it is likely that all entries are unique.

Thinking back to the sandcastle analogy in section 2.1 consider how all those bits of data (grains of sand) will all have their various levels of quality against these data quality attributes. Understanding the nature of your data quality will help reduce the likelihood that data is used incorrectly and will allow you to objectively consider what data quality issues you want to resolve.

5 The benefits of better data quality management

Typically, if we can improve the quality of data, then we will have better information to support decision making. Better decisions will lead to better outcomes/results and will in turn be likely to have better quality data arising from them. As the diagram illustrates, this becomes a virtuous circle where better data ultimately leads to better business results.



"Data quality is free. It's not a gift, but it's free. What costs money are the unquality things – all the actions that involve not getting data quality right the first time and all the actions to correct these data quality issues" (Adapted from a quotation by Philip B. Crosby)

The quotation above illustrates that the benefits of improving your data quality will be the removal of unnecessary costs arising from poor data. Better quality data can additionally present organisations with new areas to exploit.

5.1 Impacts of good and bad data

Factor	**Poor Data Quality**	**Good Data Quality**
Perceived data quality	Cost of resources to verify and "clean" data before it can be used to make it fit for purpose	Confidence that data can be utilised 'as is'
	Staff more likely to use and maintain local data sources thereby degrading the organisations data quality	Use of business systems by staff throughout the organisation.
Decision making	Poor data may give a false view of business situations, leading to poor decision making	Enabler for optimal decision making at strategic, tactical, and operational levels. Allows identification and exploitation of new products and services
Process Outcomes	Poor outcomes with lack of visibility that this is the case	Optimal process outcomes
	Poor customer perception arising from poor service outcomes and organisational reputation	Improved customer perception
Performance metrics	Effort needed to remove errors from data, and potentially manipulate data from multiple sources before performance can be understood	Easy to quickly produce accurate and trusted performance reports.
	Overall organisational performance difficult to determine	Organisational performance easily understood
Development of products and services	Difficulties in understanding current performance and trends makes it hard to identify viable future opportunities	Trusted data makes it easier to identify and exploit new opportunities

5.2 Real world examples of impacts of poor data

Poor data quality can have negative implications (financial, reputational, and other) that are sometimes highly significant. Examples include:

- A woman bled to death after a spelling mistake meant blood intended for her during an operation was sent back
- According to the UK's National Audit Office, more than three-quarters of civil service pension records (1.25m), are incomplete or incorrect, which it says has caused hardship and distress to many pensioners. The National Audit Office reported that
  - Systems capabilities not being in place to deal with processes and data on payroll and pensions
  - A lack of data governance and oversight
  - A lack of any methods to track benefits and other KPIs
- The National Health Service took the unusual step of closing down a children's heart surgery unit at a UK hospital, after data they had submitted showed that twice as many children and babies died in the unit than anywhere else in the UK. The UK media went into a frenzy; people came out of the woodwork with stories about their treatment at the hospital, neglect and near death experiences in abundance. Eleven days later and the unit reopened. It turned out that there were not twice as many people dying after all - the data that the hospital submitted to the NHS was late and incomplete; in fact, 35% of the expected data was missing completely, with catastrophic results.
- Poor data quality results in duplicate and confused patient entries on NHS systems. In other words, one patient with more than one NHS number, or the same NHS number assigned to more than one patient. The consequences can result in incorrect and mixed medical records, missed screening requests and even cancelled operations.
- The Metropolitan Police reportedly had to pay "around £1m" in pay-outs for raiding 900 incorrect addresses over a 3-year period.

- In Germany, errors in internal accounting in the nationalised Hypo Real Estate, the German National debt was overstated by €55 Billion. This was doubly embarrassing for Germany as they had previously criticised the accuracy of accounting by the Greek Government. In an era of austerity where their government has squabbled tirelessly for two years over a mooted €6-billion tax cut, Germans found it hard to fathom that their government was so suddenly and unexpectedly €55-billion better off. The net effect of the error being found and fixed is that Germany's Debt to GDP ratio will be 2.6% lower than previously thought

Germany mocked for €55-billion accounting error

NP REUTERS | October 31, 2011 2:53 PM ET
More from Reuters

German Finance Minister Wolfgang Schaeuble — Adrian Dennis/AFP/Getty Images

Like Share 0 Twitter Google+ LinkedIn Email Typo? More

By Erik Kirschbaum

BERLIN – The German government tried to deflect responsibility on Monday for a €55-billion accounting blunder that has exposed it to charges of ridicule for being inept and hypocritical after its steady criticism of Greek bookkeeping practices.

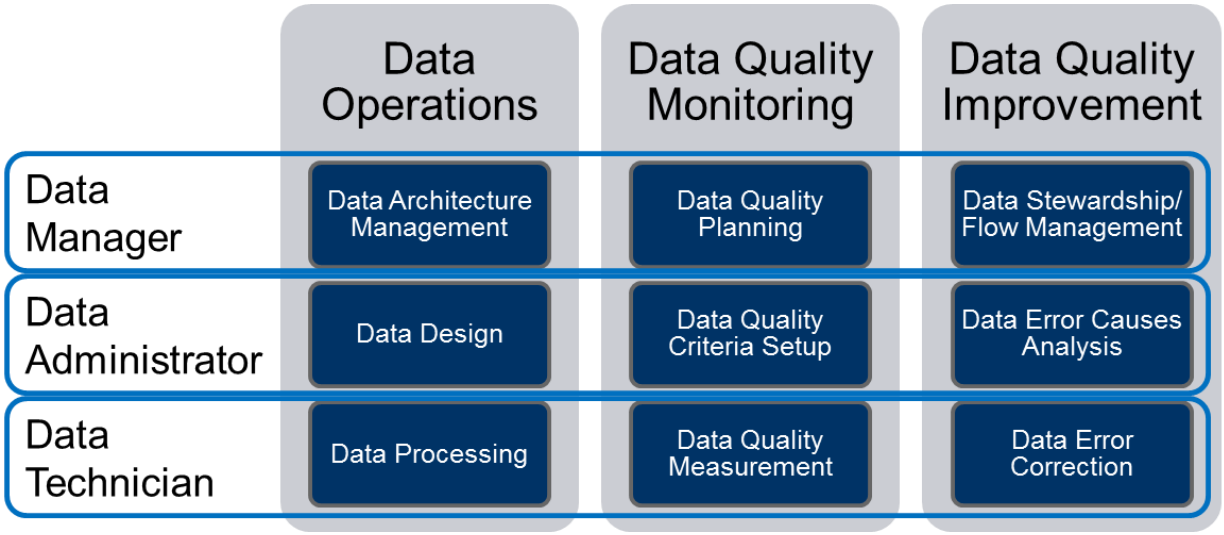
- The US Postal Service (USPS) estimated in 2013 that there were approximately 6.8 billion pieces of mail that could not be delivered as addressed. Beyond the fact that the USPS itself spent $1.5 billion to process that mail (e.g., forwarding it, returning it, disposing of it, etc.) and assuming an unrealistically low average cost of $0.50 per mailing, this is likely to result in $3.4 billion per year wasted due to incorrect address data
- At Maidstone and Tunbridge Wells NHS Trust, data on Clostridium difficile infections were found to be incomplete during the investigation into outbreaks of the infection between April 2004 and September 2006 (report published in 2007). The lack of timely and effective monitoring using complete data was one factor in delays in identifying the seriousness of C. difficile infection in the trust and responding to the situation. The report notes that "the significant outbreak in the autumn of 2005 was missed and the trust has acknowledged that it should have detected the rise in cases at that time".

The above examples all provide examples of the consequences of poor data quality. Most organisations will also have their own examples. Therefore, the benefits of improving data quality should be the removal of these negative impacts.

6 How to improve the management of Data Quality

Whilst the prospect of improving the way that data quality is managed in a large, complex, dynamic organisation may seem daunting, there are recognised methodologies that define good practice approaches that can be adopted.

ISO8000 is a set of standards relating to data quality. Part 150 of this standard describes a framework for data quality management based upon the 9-box model illustrated below.



DPA have produced a separate White Paper which explains this model in more detail and then extends it to provide a framework for data governance. This paper is available as a free download from the DPA website at http://www.dpadvantage.co.uk/data-quality-management-data-governance/.

7 Summary

This paper has shown that in the modern world data is an asset to organisations and, like any physical assets, data requires monitoring and nurturing if it is to retain its' value to the business. There are many complexities in managing your data, and these complexities only grow over time; there are potential risks if data quality management is not effective, but the benefits of getting it right are significant, even critical. The ISO8000-150 9-box model offers a framework that enable organisations to realise the benefits and mitigate the risks described in this paper and we recommend that further reading on ISO8000-150.