Simplicity of data management

Technical Note

Version: 1.0

15 May 2001

Sherringham K.

IMS Corp.
Asia - Pacific Headquarters
Suite 5 275 Maroubra Road
Maroubra
NSW 2035
Australia

Tel: +61 (0)2 9314 2908
Fax: +61 (0)2 9314 2908
Email: info@imscorp.com.au
Web: www.imscorp.com.au

Abstract

Data management is not a luxury, it is a way of lowering hidden costs, to guaranteeing quality of service and to providing customer satisfaction. Data management is not complicated, the underlying principles are consolidated data storage, accessible data about data (metadata) and storage of data in one format. Seven basic steps of data management are identified: accessions, transfer, data laundry, quality control, archiving, documentation and making it available.

1 Introduction

With developments in business and the Internet, information is becoming a ubiquity like electricity and water. To deliver cost effective valued products and services to clients, the management of information and data is becoming critical. Yet within many organisations data or information are not treated as assets. Hidden costs are continually increasing because of time spent recreating documents, finding information, reformatting and replicating the same information in different places and in different systems.

Effective and efficient data management and related policies and procedures are key to lowering hidden costs and as described below, the principles involved and implementation are basic.

2 Data – Information – Knowledge – Power

The role of information or data in business is critical. Most businesses survive using the BIDS formulae:

Brains + Infrastructure + Data = Services

Many businesses go to great lengths to get the right staff (the brains) and to ensure they are managed and rewarded correctly. The company's infrastructure (office space, desks, and computers) to do work is also treated as a key asset and is managed accordingly. By having the brains of a company use its infrastructure and its data, profits are made through the sale of it products and services.

Data are the "raw numbers" or "raw text" so to speak. Data often exists in a variety of formats in different applications and in hard copy. Information comes from a value adding process of this data. Data only becomes information when it is what a user wants and understands, Figure 1. A financial report can be produced through a value adding process but if a user does not understand it or does not want to know about it, then that report is still data. Information is data in context. The world has data overload not information overload.

From information comes knowledge and from knowledge comes power/profit/wisdom. Knowledge comes from the combination of experience, information and skills. Power then comes from the application of knowledge at the right time, in the right way. By first addressing data management, information management can be conducted, then knowledge management achieved.

Without adequately addressing data management, effective information and knowledge management is not possible.

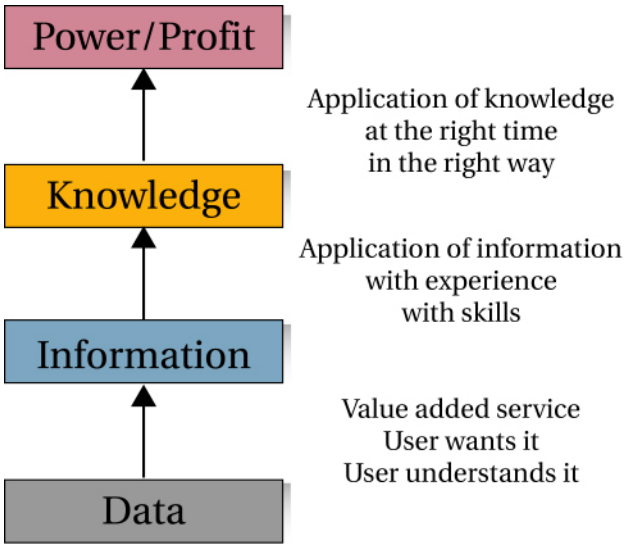

Figure 1. Conversion of data in to information, then to knowledge to realise power/profit/wisdom.

3 Basic requirement

Whilst the seven steps of data management and some basic business and operational issues are discussed subsequently, it is important to note that there are three basic requirements for data management:

1. consolidated data storage
2. accessible information about the data
3. storage of data in one format

3.1 Consolidated data storage

Storage of data in a consolidated repository is key to data management and to lowering hidden costs, Figure 2. The business environment has seen a change from the mainframe to the PC and whilst the use of distributed computing power has been a welcome development, the distributed storage of data is not. Look no further than your everyday activity. How long do you spend trying to find things? How frustrated do you get when you can't find the information your looking for? How much information is locked away on a PC that you can not get at or know nothing about? How often do you recreate things only to be told that it already exists on a colleague's computer?

3.2 Accessible information about the data

Resolving a typical business question like "what was our profit and loss for the last 2 months?" consists of the data itself, the profit and loss values, and the metadata (data about data), the time period of interest and the fact that it is a profit and loss report. One searches the metadata to find the data required. This distinction between metadata and data is important because in many data management practices metadata are not even included, where it is included the two are confused and this leads to poor search results.

Furthermore, metadata is a valuable way of trapping corporate IP because the relationship between data elements is trapped. What constitutes corporate expertise is information that is trapped in a person's head and can not be readily accessed. This is the relationships between data. In the case of a specific guest at a hotel who likes to have their champagne open when they arrive, the metadata

contain the links between the data of guest name, the champagne and the need to have it open, at a time of arrival.

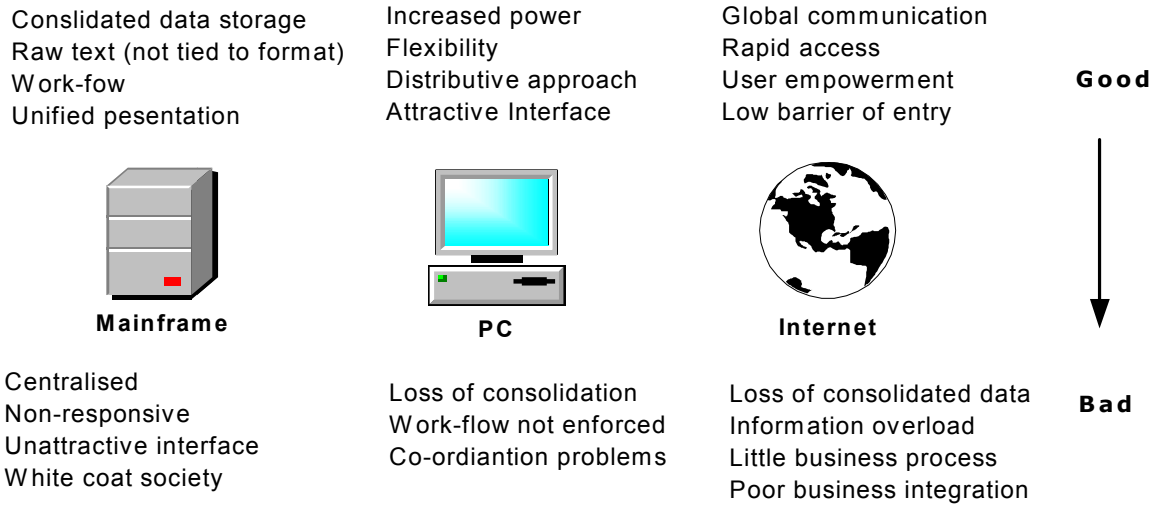

Figure 2. Loss of consolidated data storage as new technology has been implemented.

Metadata also includes important information on details like access rights, versioning, last modified date and more. All of these are critical to the smooth operation of a business. Since data is accessed through metadata, the use of a corporate metadata framework provides an effective way of linking back end systems.

The use of a corporate wide metadata framework, having metadata entered and maintained at source, by the subject matter experts, and connecting the metadata to the data is a practice that is missing in many organisations.

3.3 Storage of data in one format

Whether it is manual copy and pasting or sophisticated software, considerable effort and expense are used in reformatting data within and between organisations. By storing all of the data in "one format", considerable cost savings can be realised. In this discussion, it is necessary to distinguish between the ultimate archive or source and localised working versions and to differentiate between transactional data and documents and images.

Transactional data should exist as raw text in billing or accounting systems. By maintaining the same version of the application across an organisation, incompatibility between versions is mitigated. It is easier to transfer raw text between systems than it is to transfer data with complex formatting. All transaction data can be archived in one format, that is storage in a common database or as self-describing files, e.g. XML.

Simple document content, e.g. a 5-paragraph news story, can also be stored as raw text in either a common database or self-describing file. When it comes to large documents or those with complex formatting and images, the issue becomes more complex. With these however, the same application version and standardised software can be used to reduce the number of formats[1].

[1] With more software developers storing images and documents as XML, the one common storage format between documents and transactional data is getting closer

4 Support for different levels of technology

Not all business will be at the same level of technology as yours. That means that not only does the latest version need to be supported but that the most widely used are also supported. Like a comment, there is an initial leading edge, a nucleus where the bulk of the mass (users) is and a long tail of legacy. To support all levels of technology is not feasible and whilst it is a balance, the approach of supporting a wide a range of users as possible is good business practice.

5 The 7 basic steps of data management

5.1 The seven steps

For a simple but effective data management, 7 basic steps have been identified.

1. Accessions (tracking and knowing what you have);
2. Transfer (common format);
3. Data laundry (value adding by making the data valid);
4. Quality control;
5. Documentation (policies and procedures);
6. Archiving (banking the data);
7. Making it available.

These steps represent the overall flow, not a priority, and may not be specific to every case. In some instances, steps may be combined in to one or repeated and it may not be entirely sequential. If a simple but effective mechanism is put in place for each step, then an organisation and individuals should be able to efficiently and effectively handle their data. Out of curiosity, organisations and individuals can give themselves an honest score out of 10 for each of the 7 activities listed.

5.2 Accessions

This is knowing what data you have and what is coming into an organisation so that it can be tracked. Consider this to be equivalent to a directory of employees in a company. This directory would contain a list of employees, contact details, who they report to, where they are, what they do etc. For the data it would include an inventory of data held, who is responsible, who has access, where it is located etc.

5.3 Transfer

Data comes in to an organisation in a variety of formats and needs to converted into organisational standards for inclusion within systems. Examples of this include the manual entry of faxed data for ordering or the submission of data into a request system from an online form.

5.4 Data laundry (value adding)

This is the sorting out of what's what. This can be part of the transfer and quality control procedures and is often a very time consuming step. It is all about taking data and cleaning it up to make it useful, termed data laundry or data cleansing. An example could be an application for a driving licence where an officer would take a hastily written and incomplete application form, enter it into

the system, check the validity of the entry, verify the credentials and check for duplicate applications.

Data laundry is one of the slowest and most frustrating steps of data management. Its importance is often overlooked and the time involved is often not appreciated. Whilst data laundry is a cost, the costs associated with correcting for it later are of the order of magnitudes higher than that associated with doing it properly up front.

5.5 Quality control

One of the most important steps in the process. The consequences of having incorrect data can result in significant financial losses, e.g. a $5 million currency transfer going to the wrong account, or court cases.

For different data types and systems, organisations should have a range of quality control checks. These checks need to be monitored and performance followed. It is the quality of the hand-off between people and steps in a business process that ensures an effective and efficient business operation.

5.6 Documentation

Documentation is about trapping what has been done to the data, the policies and procedures, issues relating to the data and more. Without this documentation the data are of little use. Documentation is key to trapping corporate expertise and allows others to pick up where one stops. It also makes the whole thing a replicable process. Documentation is one of the most important steps but is often overlooked.

5.7 Archiving the data

The need for data archiving (back up and recovery) is recognised but poor policies and procedures often exist. The ability to reconstruct to a set point in time is the measure of an effective system. Banking of data is not constrained to merely electronic media but includes records management and vouchered samples.

5.8 Making it available to users

This is the final and most visible step. Whilst the rest of the system does not have to be attractive, only functional, this is where it is crucial to have the attractive interfaces. However, with out the previous 6 steps, the most attractive interface is practically useless when it turns out garbage.

5.9 An example – the phone directory

The phone directory illustrates two important points. It not only shows the importance of data management bit of the seven steps. A CEO would never allow all employees to sit there on a regular basis re-inventing localised phone lists. To prevent such a waste of resources, the data is managed and the phone directory produced and distributed. Resources are then allocated to its maintenance.

The phone directory is managed using the 7 basic steps:

- **Accessions:** The addition of new customers and the expiry of customers is tracked and managed (data entering an organisation is monitored)
- **Transfer:** All of the data are transferred into the format required for the directory - all the data put into one format.

- **Data laundry*:*** Time was spent sorting out all the data to get a product.
- **Quality Control*:*** The names, addresses and numbers were checked by some mechanism.
- **Documentation***:* The policies and procedures, quality control criteria and format specifications etc. relating to the phone directory have been documented.
- **Banking the Data:** The data for the directory are archived and a hard copy of the directory has been kept safely for future use.
- **Available to Users:** The directory has been produced and distributed to customers.

6 Policy and strategy

What is the cost to your business of sending bills to the wrong address or of under charging the client for services rendered? Data management is not a luxury, it is a way to lowering hidden costs, to guaranteeing quality of service and to providing customer satisfaction. It is not a question of the cost of data management, it is the cost of not doing it.

Where data management has been successfully implemented and lower costs realised, the following have been used:

- The mystique and complexity associated with data management has been removed.
- The need for data management is championed by the business for the business and is something that is seen to be of benefit.
- Data management is an expected activity and part of an individuals job and is part of the assessment;
- Managers are expected to oversee the management of their data and performance criteria included.
- Vendors and out-sourcers are expected to manage data as part of the contract;
- Resources have been allocated to both the setting up and the maintenance of data management systems.
- Polices, procedures and the policy decisions necessary have been implemented.
- Reporting on data management is expected, including to the board of directors and inclusion within annual reports.
- A high level vision is required and the solution is for the good of the organisation, not a high-profile specialist group.
- Standards and polices and procedures are followed. What an individual does at the desk is up to them but when using company systems, they do just that.
- Start small with something simple and move from there. Implementing simple, flexible and generic solutions have been implemented.

7 Moving forward

In any discussion on the data management issue, a tendency exists for users to look at all the old tapes that can't be read any more and the stacks of computer print outs and - justifiably - feel overwhelmed. This does **not** need to be the case. A fundamental distinction needs to be drawn between **the mountain of data that exists** and **putting in place a mechanism to manage data from this time forward:** so stopping the accumulation of un-managed data.

It is a case of "drawing a line in the sand" and managing data properly from this time forward. The mechanism can be simple, needing a small shift in political viewpoint and will take a little effort every now and then. One will then have the tools to chip away at the mountain that already exists, which will occur as priorities dictate.

8 Summary

Like other assets and resources within a business, data needs to be managed. Data management is not a luxury, it is a way to lowering hidden costs, to guaranteeing quality of service and to providing customer satisfaction. It is not a question of the cost of data management, it is the cost of not doing it. Until data are managed, knowledge management will not be achieved.

Data management is not complicated, the underlying principles are consolidated data storage, accessible data about data (metadata) and storage of data in one format. Seven basic steps of data management are identified: accessions, transfer, data laundry, quality control, archiving, documentation and making it available. Resolution of effective and efficient data management is simple but often requires a change in approach and of accountability.