

Information or Data?

by Keith Gordon

I have been involved in Data Management for eighteen years and throughout that time the question of the difference between information and data has constantly been raising its head.

The glib answer is that information is data in context, and in the past I have used that answer many times myself. But the problem with that answer is that it is too glib, and really invites the questioner to probe deeper. And if you are in presentation mode and the questioner in your audience is considerably senior to you your fumbling for an answer is really very embarrassing.

For many people it would appear that there is no difference between information and data.

Journalists will often use the two words interchangeably. An article in one national newspaper used 'data' in the headline and 'information' in the body of the article to describe the same concept. My guess is that this happened because 'data' takes up less room in a headline than 'information' and headline space was at a premium.

In the UK we have the Information Commissioner's Office. On its website it says that "The Information Commissioner's Office is the UK's independent authority set up to promote access to official information and to protect personal information". Its vehicle for promoting access to official information is the Freedom of Information Act, yet its vehicle for protecting personal information is the Data Protection Act. Is the ICO using information and data as synonyms?

When I was going through the early stages of developing the ideas for my own book with the British Computer Society, my first proposed title was "Principles of Data Management". Very rapidly I was asked if I would be prepared to change that to "Principles of Information Management". Now I fully understand where this was coming from. When speaking to people outside of the IT/IS community you will retain their attention longer if you talk about information instead of data. 'Data is technical stuff that is best left to technical people, business people are only interested in information' is the thought process (but that is a thought process with which I fundamentally disagree, but that is another issue). In the case of the book, there was a perception that a book about information and its management would sell better than a book about data and its management. I was not prepared to give up the notion that the book was about the management of data, but I recognised the need to get information into the title. So, that is how we got to "Principles of Data Management: Facilitating Information Sharing"; not a snappy title, but an accurate one.

I have already said that I am not happy with the answer that is normally used - 'information is data in context'. I am not happy with it for two reasons. Firstly, it is too glib and too imprecise; it doesn't fully answer the question. Secondly, it strikes me that it is looking at the issue from the wrong direction.

Let us look at the imprecision issue. What exactly does 'in context' mean? If I have a HOLDING table in a database with four columns, StoreCode, ProductCode, Date and Quantity, and one row in that table has the values "S36", "WGC2748", "2008-03-04", "27", do I have information or data? It looks as if I know the fact that on 4 March 2008, the store with the store code S36 held 27 of the product with the product code WGC2748. Some people would say that is definitely useful information, despite the fact that it is something we call data stored in a database. So then,

should we be changing our terminology and calling these collections of data information bases instead of databases? This is almost certainly a step too far, yet the individual data values are definitely in context when they are stored in that table. Or, does this data only become information once it is presented to an end user on a form on a screen, or on paper as part of a report?

Now the direction issue. Assume that we go along with the idea that information is data in context. This looks to me very much as if it is the IT/IS technical community, the custodians of the data, generously putting that data in context and giving it to the business people who need it as information. In other words, the technical tail is wagging the business dog. But where did that data come from in the first place. It may have been bought in (a set of names and addresses bought for marketing purposes, for instance) but it probably came from some business process that generates data, such as the reading of a barcode at an electronic point of sale terminal or an end-user entering the results of a stocktake into a personal computer hooked up to the company network.

So, what we need is a precise definition that distinguishes between information and data but that starts from the point of the view of what the business is interested in, and that is information.

Such a definition exists in international standards. It is in ISO/IEC 2382-1:1993 *Information Technology - Vocabulary - Part 1: Fundamental Terms*, which defines data as:

"a re-interpretable representation of information in a formalised manner suitable for communication, interpretation or processing."

Figure 1, which is derived from a similar figure in the standard, demonstrates this.

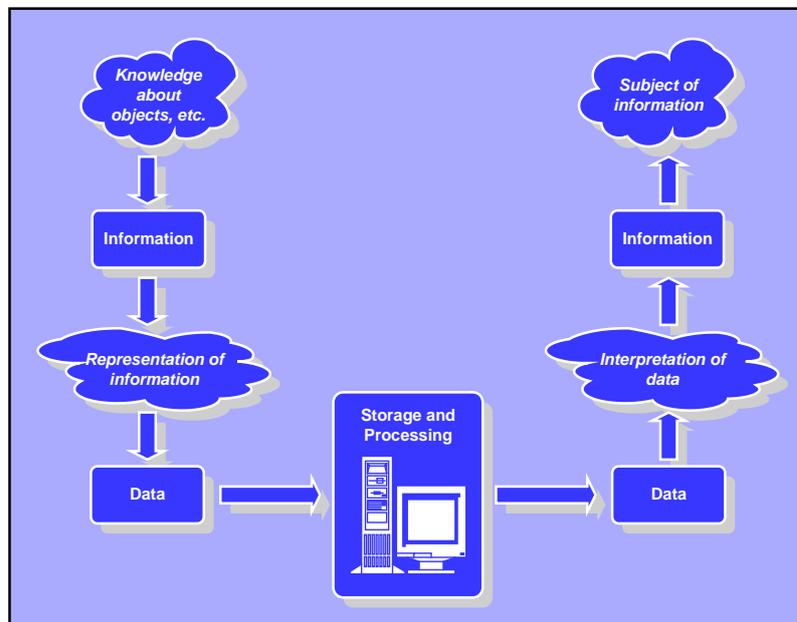


Figure 1: The relationship between information and data

The business knows something that it needs to have stored, communicated or processed. This is information. As this is put into a form where it is suitable for storing, communicating and processing (for example, it is input into a computer) it becomes data. The store codes and product codes become character strings. The date of the stocktake is in whatever date format the particular software package

uses. The quantity held becomes an integer. We have gone down the left hand side of Figure 1.

After it has been stored, communicated or processed it is still in the same or similar formats. It is still data. Only once it has been put back into a form where the business can use it does it become information again. This is going up the right hand side of Figure 1.

The trick, and this is where we data managers come in, is to ensure that the translation used when going up the right hand side matches the translation used when going down the left hand side, so that the information that comes out of the system is a true representation of the information that was put into the system. And for this we need comparable metadata.

So, I like this ISO/IEC definition of data. It provides a precise distinction between information and data whilst starting from the business perspective with information. It is now my preferred approach when asked to distinguish between information and data.

But, can we find a similar distinction between data management and information management?

One definition of data management is that used in an ISO/IEC Technical Report (ISO/IEC TR 10032:2003 *Information Technology - Reference Model of Data Management*), which sees data management from a purely technological perspective and defines data management as:

“the activities of defining, creating, storing, maintaining and providing access to data and associated processes in one or more information systems”

It then goes on to define a data management environment as:

“an abstract conceptualization of the data and associated processing elements involved in a computer system”

This is both very technical and very narrow in focus; it concentrates on the technical management of the database within a computerised information system and the software processes that interact with that database. It specifies the facilities that should be provided by a software application if that application is to be known as a database management system.

Whilst this all very sensible and laudable it totally ignores the broader issues of how you design that database so that it satisfies the information needs of the business that the information system is supporting. Before you even consider putting any data into a database you must understand the information requirements of the business and come up with a design that will enable those requirements to be met. Will it be a single database, or a federation of distributed databases? If it is distributed, how much data will be replicated? In any case, how will the database or databases be structured? How will we ensure that the data in the databases is of good quality? How will we ensure that data can be shared between different systems in the enterprise?

My view is, and has been ever since I first became involved in data management, that if it is to be worthwhile data management must start with understanding the information needs of the business. Data management is, therefore, far broader than as described in ISO/IEC 10032.

In my teaching I use a definition of data management that comes from the BCS. This defines data management as:

"a corporate service which helps with the provision of information services by controlling or co-ordinating the definitions or usage of reliable and relevant data."

This fits nicely with the distinction between information and data. It shows that data management is supporting the business ("helps with the provision of information services") and shows that we are doing that by managing "data".

The BCS then goes on to identify three distinct roles within data management: data administration, database administration and repository administration.

Data administration is concerned with mechanisms for the definition, quality control and accessibility of an organization's data.

Database administration is concerned with the management and control of the software used to access physical data.

Repository administration is concerned with the management and control of the software in which 'information about information' is stored, manipulated and defined.

But how does all of this compare with information management? Well, that depends on whose definition of information management you are using.

The former Central Computer and Telecommunications Agency (CCTA) showed the relationship between data management and the broader information management function using Figure 2.

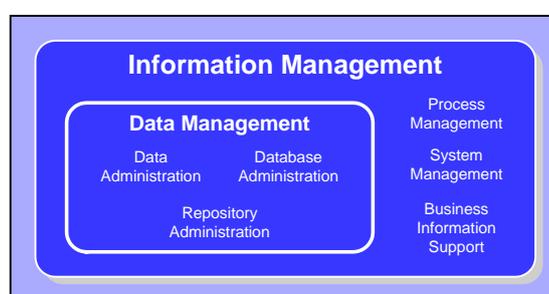


Figure 2: The relationship between information management and data management as shown by CCTA

In introducing this figure, CCTA said

"Data management itself serves the overall requirements for managing the organization's information. It is helpful to consider the wider roles and responsibilities of information management which provide a context and justification for data management."

Without trying to imply that process management, system management and business information support are trivial activities, the overall thrust of the CCTA description of this area is that data management is the major element of information management. This probably reflects the fact that the principal interests in CCTA were technological.

A contrary view is put forward by Leonard Will of Willpower Information. He describes himself as an Information Management Consultant and until January 1994 he was Head of Library and Information Services at the Science Museum in London and before that was Head of Information and Research Services in the Library of City University, of London.

In answer to the question "What is information management?" he says

"This interdisciplinary field draws on and combines skills and resources from librarianship and information science, information technology, records management, archives and general management. Its focus is information as a resource, independently of the physical form in which it occurs. Books and periodicals, data stored on local or remote computers, microforms, audio-visual media and the information in people's heads are all within its scope."

Leonard Will then goes on to say that some of the main topics information management practitioners are concerned with are those listed in Figure 3 below:

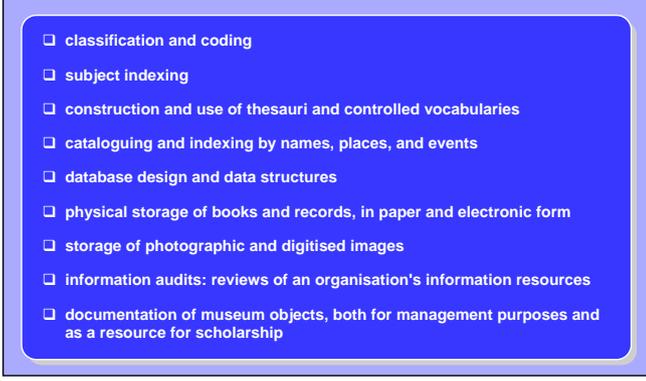
- 
- classification and coding
 - subject indexing
 - construction and use of thesauri and controlled vocabularies
 - cataloguing and indexing by names, places, and events
 - database design and data structures
 - physical storage of books and records, in paper and electronic form
 - storage of photographic and digitised images
 - information audits: reviews of an organisation's information resources
 - documentation of museum objects, both for management purposes and as a resource for scholarship

Figure 3: Information management topics as described by Leonard Will

This broader description of information management reflects the librarianship background of Leonard Will. Only two of the nine bullet points – database design and data structures; and storage of photographic and digitised images – appear to be directly 'computer-related', although computers could be used to assist with many of the other topics.

So, whilst we have a clear definition of data management from the BCS that fits with our broader concept of data management, it appears that there is not an equivalent definition of information management. What is information management tends to depend on your background, your experience and your current role. It is a term that is used by both those interested in the use of computers to support business requirements and by librarians, but both groups use the term in different ways.

My own personal view is that those of us who think of ourselves as data managers in the broadest sense - in that the starting point for our interest in data is the vitally important part it plays in the provision of timely, accurate and relevant information for business users - are as much into information management as many of those who call themselves information managers and think that data is someone else's problem. What we must not do is allow the term 'data management' to become associated solely with the technological aspects of the handling of data. Data management is an important business function and not something that should be left to technicians (and I speak as one who for many years was a technician and is proud of it).