Information Management Framework: Quality Basics

Matthew West, OBE

University of Cambridge

Abstract

Information and knowledge are important for organizations. This article discusses why information is important and how to ensure that information is fit for its purposes. It is proposed that information matters because it is used to support organizational decisions. Consequently, poor quality information can increase the risk of mistakes and disasters, and reducing mistakes directly improves productivity. The article also discusses how holding critical information as sharable, structured data is a key enabler of automation, which is another way of improving productivity. We provide ways of looking at information which are useful in assessing and understanding how to manage it, including the role of information services. Knowledge graph systems are a useful way to hold information as data so it can contribute to business and information processes. We outline the basic elements of information management including (1) Identifying information requirements, (2) The critical properties of information, (3) The information lifecycle, and (4) Applying quality management to information.

Introduction

KNOWLEDGE GRAPH SYSTEMS are a popular technology for storage and retrieval of information as data by businesses. In this article we examine the reasons for the use of knowledge graph systems and for information in businesses in general. The primary use of information in business is to support decisions so that outcomes are better, and the risk of mistakes is reduced. Despite this it is not difficult to find examples where poor quality information has led to mistakes having been made that has led to embarrassment at best and disasters at worst. The cause usually found for mistakes and disasters is that the decision was based on a lack of information, or on poor quality information, indicating that information is managed poorly, or not at all. Reducing mistakes directly improves productivity. Another important driver in improving productivity is automation – getting computers to do what people do at present. Computers are good at doing dull repetitive tasks that can be pre-determined; and coincidentally these tend to be tasks people are not so good at, which can lead to mistakes. A key requisite for automation is for information to be held as data, such as in knowledge graphs, so it can be conveniently processed. So a key part of achieving automation is not just moving data into an electronic form for people to access, but turning it into data so that computers can process it.

One of the problems that software and data can have in a business setting is the difficulty of properly accounting for their costs and benefits. For some companies, software and data are products that are sold to other companies and to individuals, in which case software and data have direct revenues. However, when a company is using software and data within its own business processes and products, it can be difficult to account for the benefits of software and data. From an accounting point of view the value of software and data is an intangible asset. An intangible asset is an asset that lacks physical or financial substance. Examples are software, patents, copyright, and trademarks, as well as data. A reduction in cost to a business, such as improving productivity or preventing disasters, will not appear as revenue on a balance sheet. As a result, software and data might only appear as costs with no corresponding benefits. Nevertheless, the costs of acquiring and managing software and data are eligible to be recognized as intangible assets provided they are (1) identifiable, (2) controlled by the organization, and (3) have future economic benefit such as revenues or decreased costs. The future economic benefits of intangible assets can be difficult to quantify because they create value in combination with other production factors. (Lev and Daum, 2004) Accordingly, it is important to specify the roles and purposes of software and data as contributors to other business processes.

Here are some simple criteria in the form of questions that will help one to determine whether information is being well managed by an organization:

1. Is there a clear understanding at the top of the organization that information needs to be managed so it supports decisions, and is there commitment to providing the resources to do so? This is the level at which it is most appropriate for an organization to identify the contribution of information to business processes so that the cost of information is properly accounted.
2. Is there a process model that identifies the decisions taken in the organization together with documentation of the information required to support them and its quality requirements? Specifying and

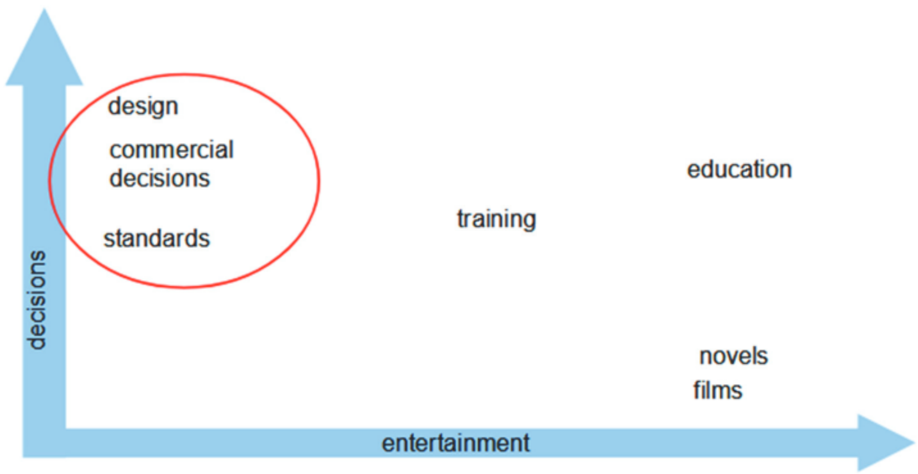
documenting process models are best done using structured information models.
3. Is there a drive to automate processes with information as structured data as a key enabler of the automation? While traditional spreadsheets and database systems can be used for process modeling, the flexibility of knowledge graphs make them well suited for the purpose of specifying process models and documenting decisions.
4. Do those who create information know the uses to which it will be put and the quality requirements of those uses? This is not only important for maintaining the quality of information but also for allocating sufficient resources and manpower.
5. Do those involved in the creation, use, and management of information have tasks and targets set for their performance of those roles? Having formally specified processes can contribute to this criterion.
6. Does shared data conform to a common data model and reference data so that data from different sources can be brought together and used without first wrangling it into a usable form? This is especially difficult for large organizations or organizations that were formed by merging other organizations.
7. Is the performance of the information management process measured so that problems with information can be identified and fixed? If performance measurements are informally specified and evaluated by people, problems can be missed or discovered too late. Automated process measurement can be more efficient and faster than manual techniques.
8. Is there an improvement process in place for the information management process so that when defects or near misses occur, the root causes are identified and corrected? This is important for all management processes.

The purpose of this article is to set out the basics of information quality management: the management of knowledge so that it is fit-for-purpose for the decisions it supports whether automated or manual, especially for knowledge graph systems. This document is aimed at all those who create, use and manage information to support decisions, so they can understand their part in ensuring decisions are supported by information that is fit-for-purpose.

Why Bother with Information?

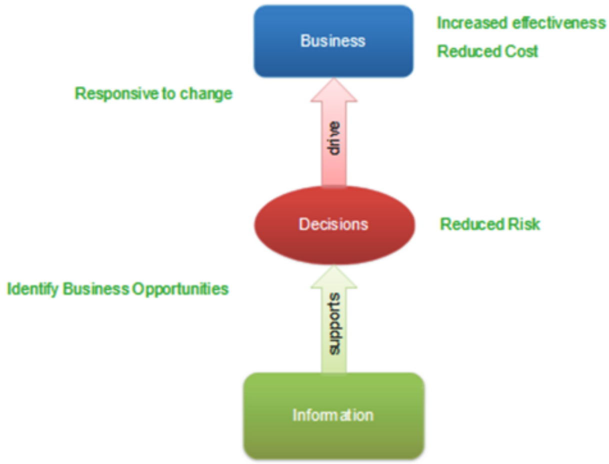
Obtaining and managing information can be expensive, especially if care is taken to ensure that the information is of high quality. Assuming that one is using information directly, and not just selling it as a product, one can classify the uses of information along two axes as in Figure 1. One axis is the degree to which the information is being used for education and entertainment, and the other axis is the degree to which the information is being used to support decisions. When information is being used for entertainment and not for decisions (*i.e.*, the lower right corner of the Figure 1), then information can be regarded as a product. When this is the case, traditional product management and accounting practices are appropriate. For other uses of information shown in Figure 1, information will be used for making decisions, at least to some degree, and it is this use of information that is the focus of this document.

Figure 1 Matrix of information uses



Information is used to support decisions which drives a business. Most obviously it helps reduce the risk of making a mistake – which reduces costs of fixing mistakes and improves effectiveness, but in doing so this also helps to identify business opportunities, and makes your organization more responsive to change, see Figure 2. When we are talking about mistakes, it

Figure 2: Why bother with information?

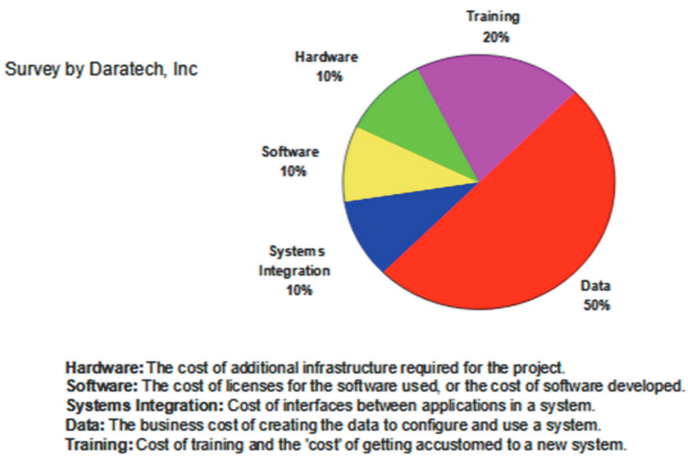


is important to understand the range this covers. A mistake might be something as small as ordering too many paper clips, where the waste from the mistake is insignificant, or it might be one of a series of mistakes that together lead to a disaster like Grenfell Tower where seventy-two lives were lost and the cost is already in the billions of dollars/euros/pounds. Eliminating waste and mistakes leads directly to improvements in productivity. For the Grenfell Tower disaster, it has already been acknowledged that a key element was the lack of a "Golden Thread" of information. (Hackitt, 2018) However, mistakes in general are not deliberate, and usually the result of the information that would have helped avoid the mistake being either unavailable or inaccurate, or being discovered too late to prevent the mistake from occurring.

As discussed in the introduction, properly accounting for the benefits of allocating resources to improving the quality and timeliness of information can be difficult. However, it is not entirely a matter of guesswork, as there is a substantial literature on risk monitoring, mitigation and management for software engineering (Pressman, 2001), which can also be applied to data engineering.

Another issue that is not widely understood is the cost of data, relative to other elements of an information system. This is largely because the cost of data is usually unaccounted for. However, in the 1990s Daratech Inc. did a survey of the costs of different elements of developing engineering design information using Computer Aided Design. Figure 3 shows the breakdown. Note that half the cost was the creation of the data itself. This comes as a surprise to many, but it should not be really, since the principal product of the design process is the design, which is information. It is just not accounted for as such. There are some unfortunate consequences of this lack of recognition, which include that much effort will be put into reducing the costs of the other elements, even when this will result in increases in costs in creating the design data itself.

Figure 3: The most expensive component of an information system is the data.



Some Ways of Looking at Information

In this section we look at some different ways of looking at and classifying information that are largely orthogonal, with each being potentially useful in particular situations.

Facts, information, knowledge, wisdom, data

In Figure 4 we show a popular distinction that is made between levels of information: facts, information, knowledge, and wisdom. Rather than give formal definitions of what these things are, we describe the usage we see made of these terms in practice.

Information

We see the term information being used with two senses:

1. As a catch-all for all the different ways of talking about, well, information including those presented here.
2. In a narrower sense as the information that informs a decision (the sense in Figure 4).

The context usually makes which sense is intended clear. Information (in the broad sense) can be structured or unstructured. Structured information is usually referred to as data (see below). Unstructured information takes the form of pictures, documents and drawings that are intended for human consumption rather than computer processing.

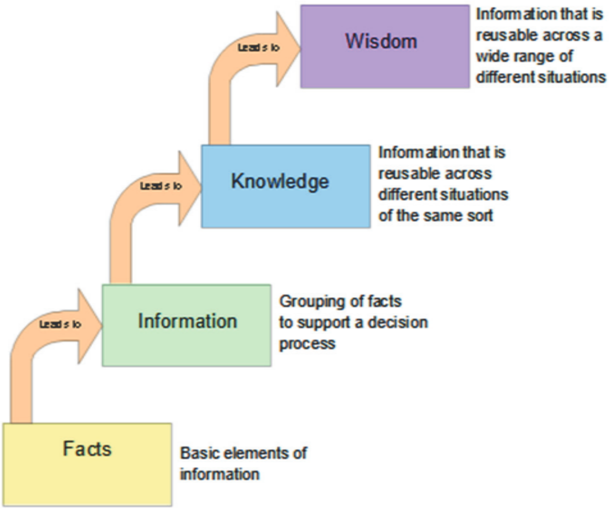
Data

Once upon a time, data (or perhaps more properly datum) was a synonym for fact. However, in recent decades it has come to mean information held in structured form, such as a database or knowledge graph, where some of the meaning of the data is determined by the structure in the form of a data model or ontology. We use that sense here. Also, we use data as both singular and plural, as most people do unless they think really hard about it. Data is relatively easy for computers to process. Computers are good at doing dull repetitive tasks that can be pre-determined, and coincidentally these tend to be tasks people are not so good at. Thus a key part of achieving automation is having information as data, that is moving beyond just holding information in an electronic form for people to access, but turning it into data in databases or knowledge graphs so that computers can process it. This is a key step in digital transformation.

Facts

A fact is the smallest piece of information, such as a triple in a knowledge graph or a sentence in a document. In many charts that show these levels, this is called data, but for us data has another more useful usage, and fact is a good alternative name.

Figure 4: Levels of information.



Knowledge

Knowledge is information that is reusable across multiple situations of the same sort. Publicly available information on the Web is an example of knowledge. Within a business, master and reference data are examples of knowledge, as are Key Performance Indicator values and designs. While any data representation technique can be used to store and access knowledge, knowledge graph systems are well suited for information that is reusable across multiple situations. (Baclawski, 2021)

Wisdom

Wisdom is knowledge of the widest applicability. It is usually about people and understanding how they will behave. Wisdom is very long lasting and not usually found in databases.

It is useful to know which level you are talking about, and when we come to consider the information lifecycle, we will look at how these interact with each other.

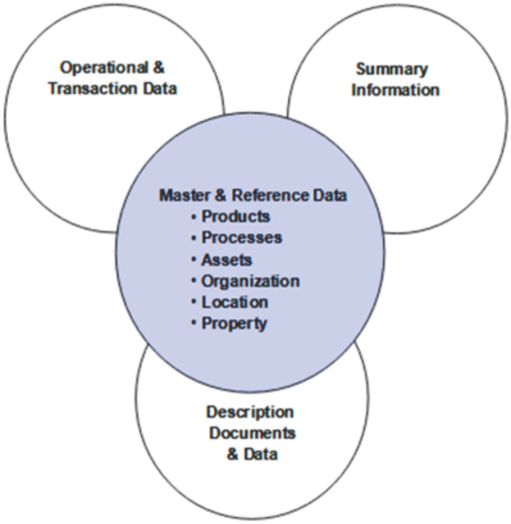
The Mickey Mouse Diagram

Another way to look at information, but in particular data, is illustrated in Figure 5, usually known as the Mickey Mouse diagram. In this view of data we have:

- Operational and transaction data, including measurements and raw transactions such as sales transactions.
- Summary information, used for management purposes and that is sliced and diced for analysis of performance.
- Master and reference data, that is used to slice and dice the summary information, and is about such things as products, processes, assets, organizations, locations and properties.
- Description documents and data, that describe what the master and reference data represent.

Generally, these different types of information have different characteristics and are managed differently and separately, held together by the master and reference data.

Figure 5: Different types of information.



The Nick Alexander Eight Box Model

Do you know what information you should have, and what information you actually have? That is the question the Eight Box model, originally developed by Nick Alexander, is designed to help. It consists of a four by two matrix shown in Figure 6, with degree of importance on one access, and whether the data changes over time or not on the other. Placing information requirements and actual records into the appropriate box tells you how you need to manage them.

Figure 6: The Nick Alexander Eight Box model, with examples.

	Unchangeable	Changeable
Category 1 Legally Mandatory	• Financial Accounts • Safety Incident Reports	• Salary Payments • Invoice Payments
Category 2 Essential to Business Operations	• Product Specifications • Production Plan	• Stock Levels
Category 3 Desirable for Business Operations	• Key Performance Indicator definitions	• Stock Forecast • Performance Measure Values
Category 4 Why keep this information?	• Design drawings for obsolete equipment	• Maintenance records for obsolete equipment

The Case of the Missing Drawing

A major oil company had a refinery on an estuary with an oil terminal on the opposite bank to the refinery and a pipeline between the two. The pipeline was built in the 1920s. In around 1990 the pipeline sprang a leak, which lead to a search for the drawing to say just where the pipeline was. No such drawing was found. The company was eventually fined £1m for, among other things, not maintaining proper records. This led to a check for what records were held to see if any other critical records were missing. Among other records they found a drawing of the kitchen table, for the canteen, that had been demolished five years previously, a poignant juxtaposition to the missing drawing. This incident led to the development of the eight-box model.

Pick a Model for a Situation

Each of the models for classifying information in this section is likely to be helpful in one situation or another when you are considering how to proceed with a particular challenge managing information. So see them as a toolkit you can dip into to pull out the right tool for the job.

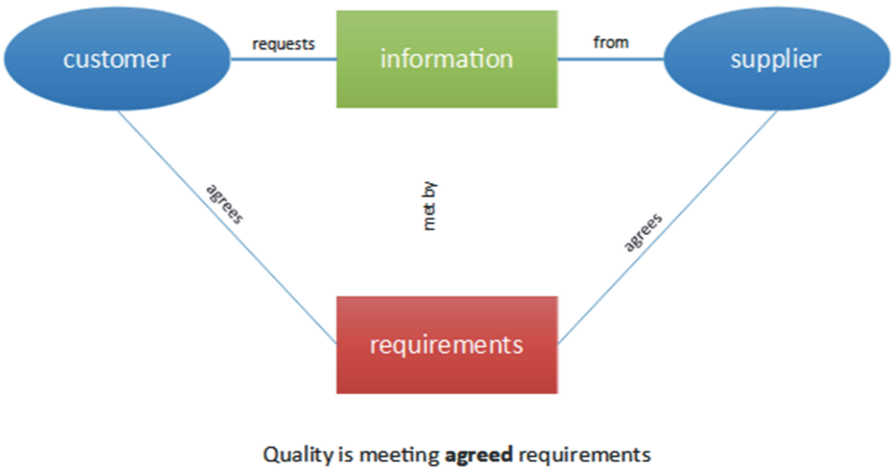
Elements of Information and Knowledge Management

The previous sections have discussed the answer to why one should devote resources to information management. We now outline how one can obtain and manage information with sufficient quality for its purpose.

Information Requirements

One can't fail to meet unstated requirements, yet so often with information, requirements go unstated, and people and organizations are surprised that their requirements are not met, often with considerable cost resulting to rectify the situation.

Figure 7: Requirements need to be agreed between customer and supplier.

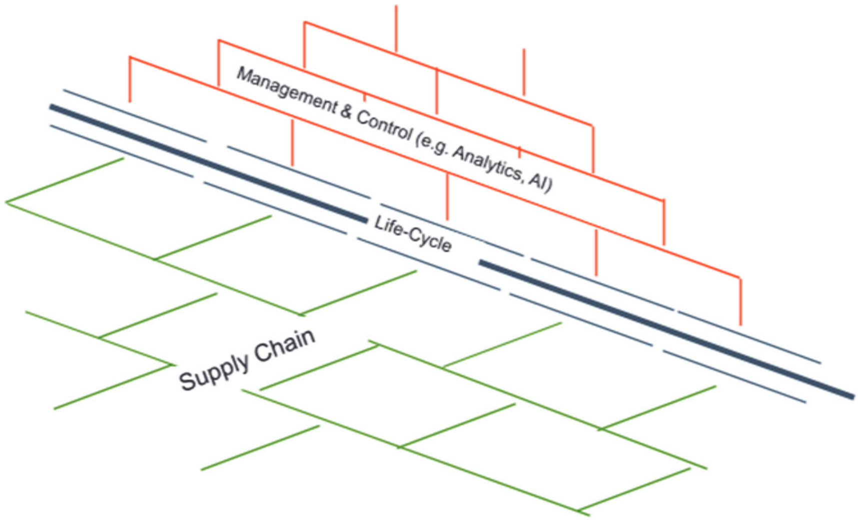


However, establishing requirements requires more than simply stating them, as illustrated in Figure 7, it needs agreement with the supplier of information; in particular about which information, to what accuracy, and when it will be delivered. There are broadly two types of information this applies to:

1. Information about a product or asset, that is essentially part of the product or asset and is necessary for its operation, maintenance, upgrade, and disposal, *e.g.*, its design and performance specification.
2. Information about operations that is used to account for and assess the performance of the business, *e.g.,* sales figures, process measurements.

Requirements for other kinds of information and knowledge such as master and reference data will be derived from these primary requirements.

Figure 8: The dimensions of business and information sharing and interoperability



Data sharing and interoperability

We need to share data when we work with others or for others, and in turn when data comes from multiple sources it must be interoperable. Data sharing and interoperability is one of the hardest things to achieve with data. Some examples of the need to share data are illustrated in Figure 8. This shows that the key dimensions to data sharing and interoperability are:

- Through the lifecycle of an asset, where data that was created, maybe a long time ago, is needed for operations and maintenance, repairs and upgrades, and finally disposal (remember the leaking pipeline).
- Through the management and control of processes where performance data is summarized for analysis across a business.

- Through the supply chain where data associated with products and services needs to be shared between organizations so that it fits with data from other sources.

The key to sharing data is achieving consistency, so that data from different sources fits together. This can be done by building a single data model based on a top-level ontology as set out in West (2011) that is used as a common language to share data.

Information Management

We are now in a position to set out what information management is. Information management is the process that delivers the right information and information that is right, to the right decision-makers, at the right time.

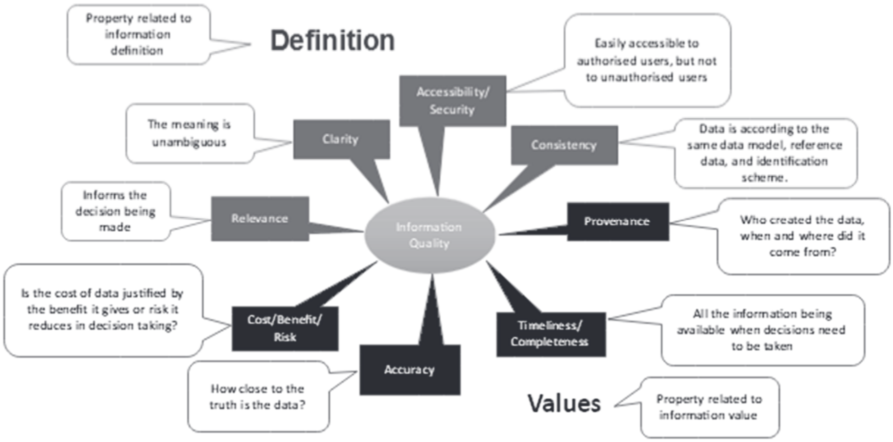
- *The right information* means that it is the information that is required to support decisions in a process, that its meaning is unambiguous, and that it is complete for the purpose and context.
- *Information that is right* means that the process for creating and maintaining the information is defined and followed so that the information is accurate, and consistent, with process performance being measured and processes improved when necessary.
- *For the right decision-makers* means that those who need to have access to the information do have access, and equally, those who should not have access to it do not have access.
- *At the right time* means that the information is available when decisions relying on it need to be taken.

For those who are familiar with quality management, you will by now have noticed that information management is essentially a quality management process. So let's look at information quality management more explicitly.

Critical properties of information

Information has many properties that might be of interest for different purposes. Here we focus on those properties that are critical to its use to inform decisions, as illustrated in Figure 9.

Figure 9: The critical properties of information for making decisions.



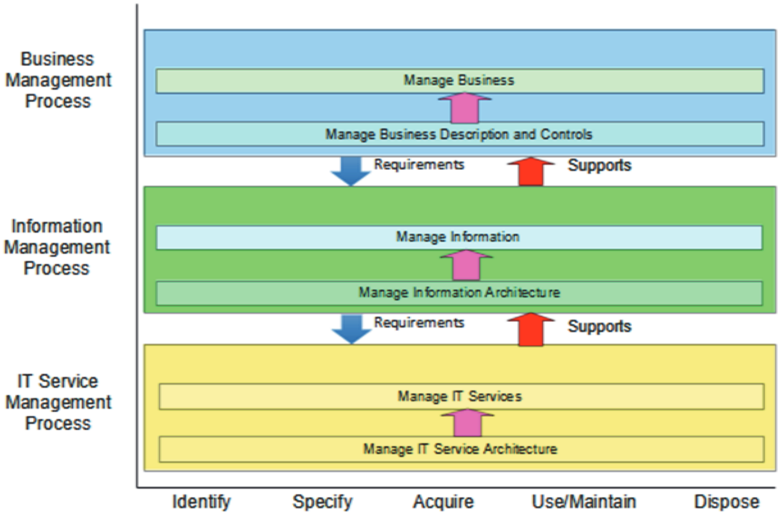
It is these properties that need to be managed. Note the division into properties related to the definition of data – that are determined when an information system is designed or improved, and properties that are about data values, and so are related to when data is created.

Where does information management sit?

It is not unusual to find that information management is missing from organizational processes. When this happens, there can be a disconnect between the business and its information because the linkage is not clear. As a result, information can be seen as a cost that has to be managed, rather than something that provides value. (Lev and Daum, 2004)

Figure 10 shows the proper relationship between the business processes, information management processes, and information service management processes.

Figure 10: The often missing information lifecycle in context.



- The *business management process* level identifies what the business needs to do to operate in terms of core and support processes, this includes identifying the information required to support decisions in the business process.
- The *information management process* takes the information requirements as input and creates and manages information to support the decisions in the business processes and identifies the requirements for the information services to provide that information.
- The *information service management process* manages services to support the information management requirements. Knowledge graph systems are increasing popular for supporting information service management processes.

Recognizing the importance of the information management process in this way means that there is a clear linkage between business and information services for meeting the information requirements, thereby having clear benefits to the business layer, and then supporting meeting those information requirements with information services giving a clear rationale for their services.

The information management lifecycle

The creation and use of information follows a lifecycle like most things, just with the added complication of different levels of information from Figure 4 to cater for. This is illustrated in Figure 11. The cycle starts with identifying requirements for information or knowledge which then gets detailed into specifying the facts that are to be collected. These facts are acquired as part of a business process, and are validated at their source before being stored. They may be used directly, for example in automation systems, where inferences may be made resulting in actions, such as if the temperature is higher than required, then heating should be reduced. Facts are collected and perhaps concentrated to acquire information which is published where it can be found for use in supporting decisions in a business process. This may cause facts to be updated. The use of information may also give rise to the acquisition of knowledge, which can also be published and used to support decisions, and be transferred and used to support decisions elsewhere as well. Part of the lifecycle management process involves archiving information and knowledge, and when it has reached the end of its useful life, disposing of it.

One useful thing one can do with the information management process is to plot popular topics. Some examples are shown in Figure 12. Although names come and go, the underlying processes were there before, and will continue after, the popular names change, which are often associated with underlying support technologies.

Figure 11: The information management lifecycle.

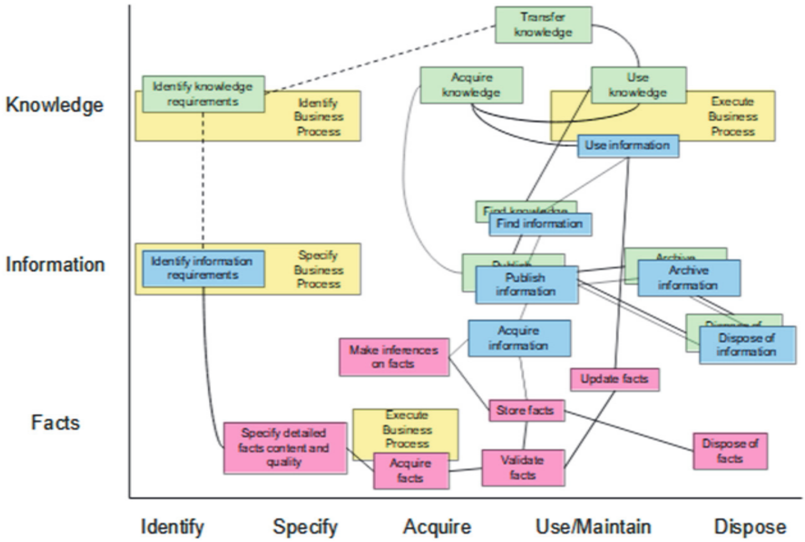
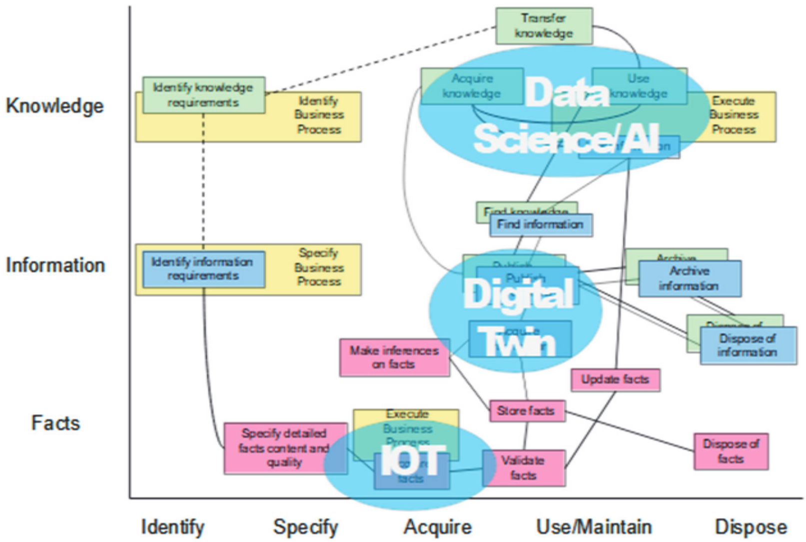


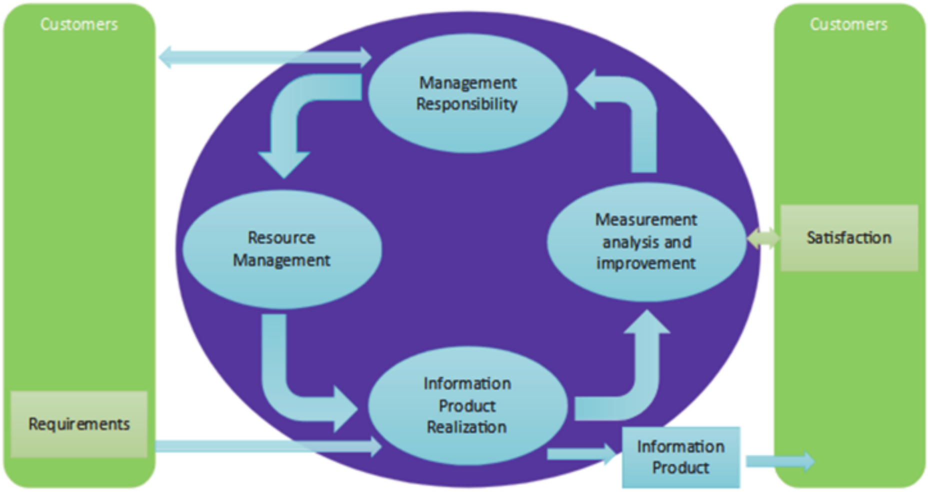
Figure 12: Some popular topics in context.



Applying quality management to the information management lifecycle

We have been talking about information quality, where quality means being fit for purpose to support decision making. However, quality does not happen by itself, it needs positive action such as the quality management process defined by ISO 9001 and illustrated in Figure 13.

Figure 13: Applying the ISO 9001 quality management process to the information management lifecycle.



The first thing to note is that the only change I have made to this from the original at this level is to substitute "information product" for "product". The second thing to note is that the information management lifecycle from Figure 11 is the information product realization process in Figure 13. This is the process that quality management needs to be applied to.

In some places quality management has turned into a tick box exercise with a poor reputation, and I want to emphasize that is not what is involved here. Quality management needs to be a cultural change from a blame culture, where when something goes wrong you look for whose fault it was, to an improvement culture, where when something goes wrong or there is a near miss, it is seen as an opportunity to improve your processes to reduce the chance of recurrence.

If we start at the top in Figure 13 and as broken down in Figure 14, with management responsibility, then this is recognition by the C-suite that information is critical to business success and there is:

- Commitment to managing information and what is needed to achieve that,
- A focus on meeting the needs of the customers for information – those taking decisions,
- Development of corporate policy for information and information management,
- Establishing the responsibilities for data, including setting tasks and targets that can be delegated down the line to those who create and use data, and communicating throughout the organization what is expected,
- Management oversight of information management processes,
- Regular management review of information management progress and performance.

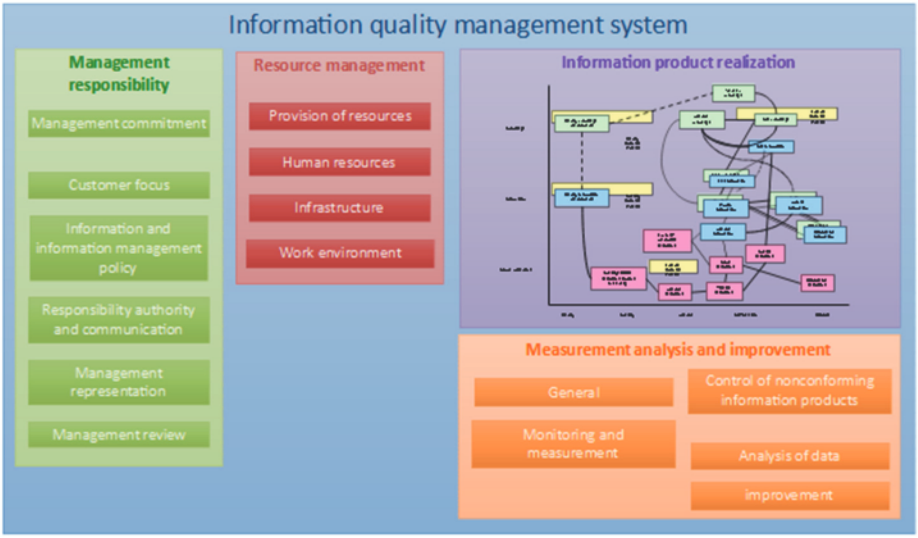
The next key element is providing the resources to do the job. It is surprising how often information management is added as a task to someone who is already fully loaded, and then without tasks and targets being set. It is not hard to see how that will go. Resource management has a number of elements: Human resources, Infrastructure, and Work environment.

As already mentioned, the information product realization process is the information management lifecycle process already described. The key things to note in addition are ensuring the quality requirements of the customers for the information are understood with processes established to ensure they are met, and to check that customers are satisfied with the information product supplied.

Finally, we have the measure, analysis and improvement process. This includes:

- Monitoring and measurement of the information product to ensure that it meets requirements,
- Control of non-conforming information products, by *e.g.,* cleansing,
- Analysis of non-conforming products to identify the root cause of the non-conformity,
- Making improvements to processes to ensure the non-conformities do not recur.

Figure 14: Information quality management process detail



Establishing an improvement process around your information gives you a way to pull yourself up by your boot laces, systematically improving your performance in information management, and as a result, improving the quality of decision making in your organization.

Conclusion

In this article we have:

1. Shown that information matters because it is used to support taking decisions, so that poor quality information increases the risk of mistakes and disasters, where reducing mistakes increases productivity,
2. Identified that a key enabler of automation is holding information and knowledge as data that can be managed and accessed via services such as those provided by a knowledge graph system.
3. Provided some ways of looking at information that are useful in assessing it and understanding how to manage it,
4. Outlined the elements of information management including,
    - Identifying information requirements,
    - The critical properties of information,
    - The information lifecycle, and
    - Applying quality management to the information lifecycle.

Acknowledgement

This paper is based on work originally done when the author was working for Royal Dutch Shell. This has more recently been adapted and improved through work for the UK Centre for the Protection of National Infrastructure for a forthcoming report “Information Quality Basics”.

References

Baclawski, K. (2021) "A Knowledge Graph Data Model and Query Language", In *J. Wash. Acad. Sci.*, to appear.

Hackitt, J. (2018) "Building a Safer Future", Retrieved on 6 September 2021 from https://bit.ly/3tnXHu8

Lev, B. and Daum, J.H. (2004), "The dominance of intangible assets: consequences for enterprise management and corporate reporting", Measuring Business Excellence, Vol. 8 No. 1, pp. 6-17. https://doi.org/10.1108/13683040410524694

Pressman, R. (2001) Software Engineering: A Practitioner's Approach, Fifth Edition, McGraw-Hill, Retrieved on 6 September 2021 from https://bit.ly/3neJ37d

West, Matthew (2011) Developing High Quality Data Models, Morgan Kaufmann