

What is Information Management?

Information management is what ensures an enterprise and its data to not part company. As enterprises have adopted computers, the server and the network have replaced the filing cabinet, Xerox machine and internal post. The ease with which information now moves has created new opportunities and ways of working but with these come new pitfalls. Too many enterprises are making major investments that are stuck in out-dated and inappropriate ways of managing information - they are sucking cost into the business under the illusion that they are making it more efficient.

Information management links people to the information they need to do their job. It is not about filing systems or efficient use of computer resources - that is *data management*. It works with *sustainment management* to keep information is available for as long as it is useful in spite of changing computers and software. And where people are co-operating, they need *configuration management* to keep track of working drafts and *information management* to share the finished article.

However, bitter experience shows that most "Information Management Solutions" are only data management systems with a few add-ons for co-operative filing. They may be sold as "out-of-the-box solutions", but enterprises succeed by being good at their own unique job and not by simply adopting tools and techniques of another. Information management is about how *your* information informs *your* people in *your* enterprise; it is not about having a single, very large electronic filing cabinet.

How Big is your Information Management Problem?

We all use information management. At home, I stick bills on the fridge to remind me to pay them - this sort of information management system solves only a small problem.

A bigger problem is the number of books I and my wife have. Much of the non-fiction sits on three large bookcases in the study, mostly sorted by subject, but with a separate shelf at the bottom for big books and a further set of shelves on the wall for my computing books. Fiction lives in the sitting room, except for lying-in-bed-detective-novels that have their own bookcase upstairs. Cookery and gardening live in the kitchen. Our broad information management strategy is to store books near to where they are used, and that gives us a clue as to where to start looking. Finer level management is by topic and shelf, so to find a book on a particular topics we need to go to the right shelf and read along some twenty or so spines; aesthetics (and it being too much effort) mean that individual shelves are not further sorted.

Information management is about matching people to the information they need. One crude measure of "bigness" is the number of pieces of information to be managed. Computers are very good at big, crude problems, like sorting a phone list into alphabetical order. However more complex problems require more complex approaches: finding a book on a particular topic is more complicated than an alphabetical list of authors or titles. A more important measure of "bigness" is diversity: the more diverse the groups of people, the topics and the types of information, the bigger and more complex is the information management problem.

At this point somebody usually shouts "metadata", as if they have just won at Bingo. Metadata is data used to help manage a problem, but it only as good as the people who create it. If the information is used by a diverse group of people, each using a variety of different sources, then the

person creating the metadata needs to understand those people and the way they think about what they do. To take an obscure example, when researching the "Bournemouth Motor Rail Service", will the person be looking for information about Bournemouth or about railways? In fact, the topic is covered in the Wikipedia entry "Meyrick Park Halt railway station" and the article is found through full text search for "Meyrick" - and I knew where to look because my father wrote the pamphlet: social networks are often a key factor in connecting people to information. Understanding who the potential audiences are and how they look for information is a complex and time consuming problem.

There are other important measures of "bigness". However, the main point is that any requirement for a system that "provides users with the information they require" without a comprehensive view on the size of the problem amounts to nothing more than shouting "HELP!"

Complex Data or Complex People?

Information management is rather more complex than simplistic analogies like "desk top" and "filing cabinet" claim. This doesn't mean that the user needs to see the complexities - they should be hidden below the user interface. But what goes on under the interface must be as complex as the way information is handled by the enterprise - but no more complex. And that requires clear thinking about what is going on underneath.

The starting point is the "unit of information". I have seen approaches which start by declaring "everything is a document", which in reality means that the system is based on an undifferentiated data goo. It is possible to build an enterprise round the 20th century model of how things were done on paper, but that enterprise is unlikely to be sustainable far into the 21st century. Databases, wikis, webs can only be forced to fit in with "one size fits all" if half their functionality sits outside the "Information Management System" - i.e. the system doesn't really manage the information.

Consider, for example, writing a complex specification for an aircraft or business service is usually the work of several hands, each contributing their own chapter or section. Historically (1970's), each section of the document was individually typed, and the final document released only when every contribution was in: only then could downstream work begin. The *document* was the old unit of information. Now it is possible for the specification to be developed in a shared environment, with every section (the *section* being the unit of information) managed for both maturity and visibility downstream. This allows downstream preparatory work to begin early, so enabling immediate feedback on ambiguities and potential problems. The result is a higher quality output and a reduced development lifecycle. The unit of information is defined both by what makes a coherent package and whether that package can be effectively managed.

This complexity occurs because any individual has only a limited knowledge of the technicalities of each problem. The mathematician working on the signal processing algorithm has a different way of thinking to the electronics engineer making the circuits involved and they have a different perspective from the designer trying to ensure that enough air flows to stop the electronics overheating. In a bank, foreign exchange is a different expertise to investment management. The hard part of managing the product is to ensure that you have the people with the right knowledge available to do every part of the job: the complexity of information management is ensuring that everyone knows enough about what everyone else is doing to not trip them up.

Literature provides few examples of systemic complexity. Farce exploits the collisions when several characters each searched their own interpretation (or misinterpretation) of the implementation.

It is the job of configuration management to say "Calm down" and ensure everyone works from the same version of the truth. Information management looks at a bigger canvas, like that of "Game of Thrones". By the end of series six we are pretty sure that Daenerys Targaryen and her dragons will be key to defeating the White Walkers, even though she's never heard of them, spent six series on a different continent and has only got two more series to get there. As viewers, the complex interlocking stories of Game of Thrones are made easy to follow by setting each thread in a well differentiated mini-world, where the stocktaking architecture, the clothes and the accent provide the clues as to which story is currently in play. Big corporations provide much blander contexts, so that the complexities are less evident. Sometimes these complexities become evident in failure, as when an administrator is replaced by a set of Internet forms, but then the day comes when nobody (except the now redundant administrator) knows which form is used for a particular once-a-year circumstance, let alone what the form is called or where it is stored.

Complexity occurs when no single person has the whole picture. However, complexity does not grow in proportion to the size of the problem, but has step changes as it passes certain thresholds. For example, when going from a single team to a team-of-teams (say from five to fifteen people), communication needs team leader to meet to make sure everyone keeps in step. Growing to a team-of-teams-of-teams (about 50 to 80 people) suddenly requires more formal controls - typically simple configuration management, but not the heavyweight process needed when the project staff grows to thousands. Conversely, having a large organization does not necessarily mean the processes are complex: for example, a large retail outlet may have thousands of staff, but most do a very similar job of selling and stocktaking - complexity is mostly confined to a relatively few head office staff who deal with specialisms such as finance and logistics.

Information Management in Context

As hinted above, Information Management is best treated as one of four *specific* topics that get grouped under the vague topic of "information management". They are (figure 1):

- *Information management* - a people-facing topic that deals with how to link people to the information they need;
- *Data management* - a computer-facing topic covering the physical movement and storage of data within a computer;
- *Configuration management* - dealing primarily with the development phase of complex "data items", making sure that the right versions come together in the final output;
- *Sustainment management* - ensuring that information remains usable when it outlives the software and computers that created it.

It is a feature of management that it does very little itself, but co-ordinates a variety of specialist workers to actually get the job done. The current trend in software is to implement these "workers" as services. The differences between the four "information management" topics can be seen in their different usage of common services, as illustrated below with four generic service collections:

- Physical Location Management;
- Access management;
- Search management;
- Trust Management.

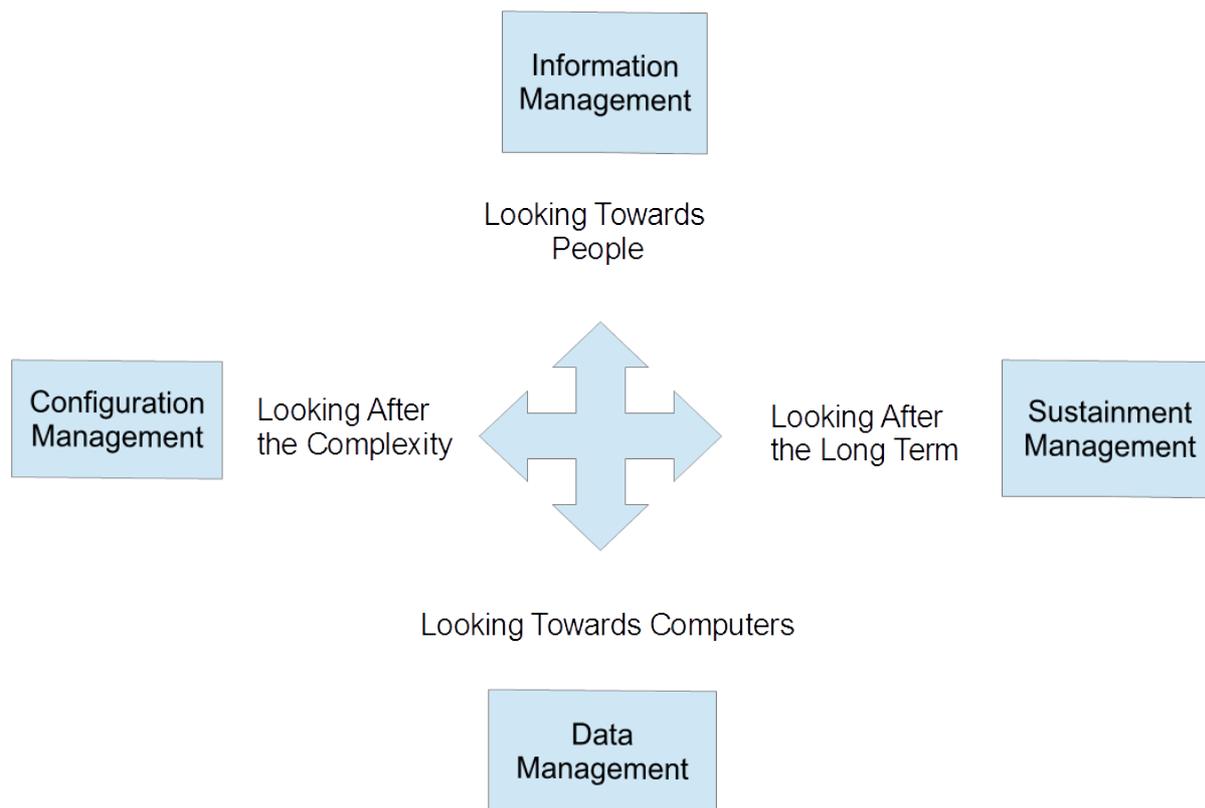


Figure 1: Information Management in Context

Physical Location Management

Physical Location Management finds *this file on this storage device*. Physical Location Management is one of the key functions of Data Management. For example, as storage devices break down or wear out, the data on them is migrated to new devices and that the paths to the data are updated. I know a manager of a high value data collection who checks every day that the list of items in his collection was complete and correct, mainly because his IT service provider had been known to replace disc drives without updating the physical access paths, creating holes in the collection. If these holes are spotted quickly, there is much more chance that the error can be traced and rectified.

Information Management does not care where the data is physically located, but it does care who wants to access that data. If the users want a high volume of data with a high availability, then Information Management identifies this a "service level", and tasks Data Management with the job of creating a mirror site that can provide that level of service. In most organisations, the users and the service levels are fairly static, but in some organisation such as the military these service levels can change quite dramatically. There is a separate replication requirement for Sustainability Management, which requires data is stored at multiple sites in case of disaster and requires diverse storage devices to mitigate risks of storage technologies hitting a dead end.

Access Management

Information Management identifies the conditions under which users can access may access information - which might include confidentiality, privacy or intellectual property - and links the access conditions to the users and any obligations that access may entail. Access Management provides the various mechanisms that enforce the conditions. For example, corporate users might access a copyright library under a corporate licence while others may may have to pay per view.

Access management typically sits between Information Management and Data Management, telling the Data Management services whether they can supply the particular file to the user. Configuration Management may use Access management to ensure that one component supplier cannot see the design from a rival component supplier. Sustainment Management relates the information back to the different Information Management policies as they have changed over time so that, for example, if one supplier takes over another, they are now allowed to see historical information that was forbidden when they were competitors.

Search Management

Search Management enables people to find the information they are looking for. Search works well when it finds the things seached for and eliminates the irrelevant; it works badly when it misses the important or picks out too many irrelevancies. Information management identifies the user communities and the search terms they employ. Data Management speeds up search by pre-searching and associating search metadata with the files or by building common indexes. Sustainment management identifies how information needs to be re-indexed as the user community changes its search terms over time.

Trust Management

Trust is key to configuration management, since downstream users are concerned whether they are looking at the latest version of data and whether that data is mature. Maturity is a measure of how much information is likely to change and the scale of impact of that change: a document awaiting only the spelling checker is mature, whereas a contract where the price is "to be specified" is immature, even if every other detail is complete. Trust Management provides the services that help record and even measure maturity. Trust also is used by Sustainability Management to show that historic data is authentic, and particularly that it has the evidential weight need to allow it to be used in legal cases. And it provides Information Management users with confidence that they have the right information and all the right information.

Summary

The vague topic of "information management" is a complex nexus of various overlapping problems. I have broken it out into the four distinct topics of Information, Sustainment, Configuration and Data Management, each with their own scope and focus. By doing this explicitly, one can then start to manage "information management" by separating out distinct problem areas and by measuring the scale of each problem.

If the enterprise only keeps its data for a couple of years, then sustainment management is probably not needed, except for contracts, tax, pensions, health and safety records, etc. (not to mention disaster recovery). If the enterprise only writes documents, with only two or three people contributing, then Open Office and e-mail are as much as you need for configuration management (as long as you don't share spreadsheets or have a web site). If you only have a single office of half a dozen people, then you can skimp on the information management (as long as people don't leave or go on holiday). And Data Management is solved by the computer directory structure - as long as

you have only one computer and print everything out in case the computer is stolen. In practice, smaller enterprises can muddle through as long as they have the tools to deal with their biggest problem.

However you may be working at the large scale - say 5,000 skilled employees involved in shared development of multiple lines of complex products which lasts more than seven years (a couple of generations of computer). Here the enterprise should separate out the different areas of information management and assess the size of each problem. And unless it does this, there is a substantial risk that when an enterprise buys an "information management solution" that it buys a system that solves a different problem to the one it has, and in doing so, leaves its users to fight the new system to deal with the actual problems they face.

Or, more likely, you sit somewhere in between the extremes, but without proper measures of your information needs you are in danger of buying a system that is too complicated (and expensive) for most of you want but doesn't solve the one complex problem you actually have.

Or, perhaps you are happy that your "information management strategy" really just says "HELP!"