The challenge of 'e-Documents'

The world of business is moving increasingly towards storing and exchanging documentation almost exclusively in electronic formats. While this process has many advantages, a major report we have been working on, based on a data set of more than 250,000 records has highlighted some challenging problems. As it becomes increasingly easier to trap every iteration of a document finding useful information becomes harder in the thousands of records stored across multiple systems; determining which out of 6 different copies of a document is the 'original' and checking to see if any other five have different information that may be relevant becomes almost impossible. We were finding one starting email generating several different 'email trails' with different people responding to different 'trails' without necessarily knowing (or bothering with) information in another response contained in another 'trail'.

Access to 'everything' is useful if you know precisely what you are looking for, but is far from helpful if you are trying to discover what happened and don't have specific questions or points of reference yet. The e-Document challenge is to build systems that support both needs plus a third:

1. Specific queries such as 'show me everything related to xxx' (eg, order number ACd-123579). The query should deliver every document that has the specific search item included within its information set.
2. General queries such as 'show me everything related to cranes in March 2020'; with options to progressively refine the query.
3. Business Intelligence (BI) analysis looking for trends and emerging opportunities in the accumulated data.

Answering any of the queries above requires access to stored documents in an integrated or linked storage and retrieval system. While this is a rapidly evolving area, there are two basic types of document storage and retrieval systems:

1. Systems that rely on tag-lines or document characteristics for sorting and searching (for example, document titles, tag lines in emails, dates, and authors, senders, receivers, etc.
2. Systems that allow the full content of most document types to be searched (think Google).

Many of the more sophisticated document management systems support both of the above processes.

The next consideration is how information is captured. Some systems simply store information either through uploads or routine trawling of other storage systems (eg, all of the company's emails are uploaded to the document retrieval system each night). Other systems manage the information allowing version control, etc and originate the sending of information – work is done using the system. The latter option is typical for many team collaboration tools, BIM[1] and other design management software.

[1] ***Building Information Management (BIM),*** see: https://mosaicprojects.com.au/PMKI-XTR-005.php#Process1

If your organisation is not using one or more of these systems, it soon will be! Document management tools solve many problems typically found in paper-based systems but introduce a suite of new issues and challenges. The focus of this article is highlighting a few of the more important factors needed in an efficient system:

- **Create one source of the truth**. As people become more used to the 'system' they begin to rely on it and if something is not uploaded, stored or created in the tool, it ceases to exist. Discipline and processes are needed to make sure a document retrieval system contains all of the documents; you cannot rely on people remembering to do the right thing and if someone is doing something unethical, they will try to evade the system. Developing robust systems for authentication, verification and 'signing' documents are also required, in particular ensuring a 'signed' document cannot be changed later. The solution is system design and automation, but this is not easy.
- **One document, one record**. Send an email to 10 other people in the organization and you immediately have 11 versions of the one document scattered across various email accounts (this is before 'reply all' and email trails start to build). Your document management system needs to be smart enough to recognise identical versions of the same document keep the original, and archive the 10 copies. However, when someone changes the email (maybe by forwarding it), you have a new document, and the process gets more complex if there are attachments. The solution to this is a system that can manage families of documents.
- **Finding something.** This is the biggest challenge with massive archives of documents (and was central to our work of the last few months) how do you find information? A search based on document contents may seem the best option, the google search engine and our cloud-based Microsoft system search like this. But if you Google *PMI PMP exam change,* you get 891,000 results. Google's systems decide which of the pages it will show you and the sort order. But if you are looking for something specific there's a lot of work to look through the list. This gets even more difficult if you want to check if something did not get documented, a null-result may mean the alleged document does not exist, or it may mean your search terms are slightly out.

Developing systems that balance providing information that you need against burying you under masses of information requires the wisdom of Solomon. Artificial intelligence can help if the search is routine, but for an important ad hoc search you are probably on your own.

One way to help focus searches is by structuring the information, using folders or codes. The problems are minimising misplaced information (wrong folder or code) and persuading everyone to use the system. Again, system design is central to developing processes that work. This type of categorisation can be linked into an effective knowledge management (KM) system, but KM is at another level and involves transforming the information held in documents into accessible knowledge[2].

The concept of a paperless project has been around for a while now and electronic document management systems are becoming increasingly common, many are really good at helping teamwork and collaboration on small projects. The challenge now is to scale this concept up to the enterprise level and develop tools that can quickly provide you with the information you need from a pool of several million documents.

---

[2] For more on ***knowledge management*** see: https://mosaicprojects.com.au/Mag_Articles/SA1065_Knowledge_Management_is_more_than_Information_Management.pdf

First Published 21st October 2021