



# Search and deploy

**Whether it's libraries, business or education, in today's world it's not just what you know but how you use it that really counts**

By Elspeth Hyams

"Although the world is awash with data, that is not the problem professions face. It is the lack of good filters." So says the latest annual report from information specialist Thomson Reuters. In fact, when it last counted, there were already 800 exabytes of data in the world – that's 800 billion gigabytes.

Most of us do not want all of the data, of course – we just want the right bit. Not finding what we need wastes lots of time and is a significant hidden cost in business. The larger the organisation, the more time saved, and the greater the impact on the bottom line.

Intranet technology, content and knowledge-management systems, corporate portals and workflow solutions have improved the lot of the office-bound knowledge worker greatly

*Not finding the data we need wastes lots of time and is a significant hidden cost in business*

in the past few years. But the amount of material everyone has to deal with, even with effective filters, has led to a very human response: when in doubt, ask your real-world friends. Almost as significant as collaboration, technology has been the cultural impact of social media, particularly the Facebook effect.

In many organisations where internal knowledge is an important business resource, there is implementation of Microsoft's SharePoint. This platform is now



used in some form by 90 per cent of organisations, according to content technology analyst Real Story Group.

But the corporate world has reservations about SharePoint's robustness. It is thought to be best suited to organisations that need to make communication and sharing easy – such as a charity like Christian Aid, which

needs to co-ordinate its relief effort for disasters of the magnitude of the Haitian earthquake.

In environments where there are strict regulatory requirements or where client confidentiality needs to be respected, information management gets more complicated. Workflow and systems architecture ideally need to capture and retrieve anything that might be needed in a legal situation requiring full disclosure – and that is technically much more challenging.

Even so, the Facebook effect has been so influential that this year technology analyst Gartner is recommending that enterprise content management systems more than five years old are replaced by ones organised around processes, social-content management and integration.

For search tools on the web, keyword favourites such as Google are falling out of favour. There are other alternatives that include specialist and multi-source engines – more than 170, according to independent expert Phil Bradley.

With a database product, though, users can benefit from a publisher's "discovery" platform, which provides a Google-like search for highly structured, indexed content in a controlled environment.

Unfortunately, companies' content is not



usually indexed at all.

Certainly, enterprise search tools are big business, and the best ones are being snapped up by the IT giants. Microsoft long ago bought Fast, but Oracle has just bought Endeca and Hewlett Packard is acquiring Autonomy, one of the most successful ever British software firms.

But the effectiveness of search still depends on how well the search engine is integrated with corporate systems and the in-house taxonomy. It's never a plug-and-play solution.

New technology like this can support process efficiencies but if it is to live up to expectations, training in good information management practice is essential. A technology implementation in the information world needs as much attention to change management as other IT projects.

## Education and research

Our ability to search academic journals efficiently has speeded up the research process itself – thanks to the work of publishers. Not only can we mine the structured and hyperlinked evidence base, there is more and more material available – millions of articles now.

A paper only gets published if it has been through peer review. Universities and funders value this objective quality assurance process, and all the other aspects of publishers' service (such as reference linking and metadata enhancement). But they do not like being prevented by licence restrictions from freely sharing papers with their peers.

In recent years there has been a backlash from the UK's funding councils, which want research results disseminated widely, the so-called open access movement. But publishers fund their services by selling subscriptions. They need to protect their intellectual property if they are to do so.

Most publishers are accepting that change is necessary and are experimenting with open access publication. Even so, still only 7.7 per cent of academic journal articles originating in the UK (which produces about eight per cent of total global scholarly output) are published in this way.

Recently there have been stand-offs on journals prices, with some prestigious university libraries in the US and Britain threatening to cancel their subscriptions to journal packages

unless they get a better deal.

Meanwhile, companies such as the publisher of Nature are experimenting with open platforms that publish reports on research in progress. It all involves cultural change – and it is slow as attitudes among researchers themselves are ambivalent.

Much content from publishers is still only available as PDFs. It's the fear of piracy that keeps it this way, because PDFs can't be manipulated or mined for text and data. But in the textbook market, publishers are being more daring.



## A technology implementation in the information world needs as much attention to change management as other IT projects

Educational publishers such as Pearson and Cengage Learning are breaking textbook content up into smaller units which can be rented rather than sold, or shared and used in different contexts.

They are also developing or acquiring platforms and learning environments of their own. Their business models are evolving to more closely reflect how students want to study and engage with their peers. It's all part of the shift towards student-centred learning.

In the long term, this new delivery model raises big questions about the management of institutions' learning platforms, student portfolios and institutional records. Who is responsible for this content? Who manages and curates it? Could there be process efficiencies from integration? And who will do these jobs in future?

## Libraries and preservation

Key issues for companies, academia, government and so-called memory institutions – libraries, archives and museums – are selection, capture, preservation and making resources findable and useful. They may also need to make resources available in new ways: images, music, video and sound.

National libraries and archives are heavily involved in digitisation and preservation projects, in partnership with innovative technology companies – with Tessella's digital archiving software at the National Archives among the examples.

Also at issue for information professionals is material closer to home, in their own organisations. How should they identify it, who should evaluate it and decide what to keep? How much of the process can be automated? Who should describe it or will top metadata and text-mining do the job?

In pharmaceutical companies, legal or health practices, construction or other regulatory environments, preservation requirements are clear. But in scientific or social science research, researchers use only about 10 per cent of the data they generate – so what should they keep? There is a need for curation services of many kinds, some of them yet to be invented.

Academia is working hard to identify and curate what are known as open-learning objects which can be reused, and in general, digital asset management has become a new competence.

The capabilities of technology are already far ahead of how we are currently rethinking what we might do to organise and exploit information better. It's exciting or scary, depending on your attitude to change. Technology can, but need not be, expensive so decisions about whether to invest in new tools or information products need a context.

It's no longer enough to be an information, technology or business specialist. The days of a department operating in its own silo are long gone. Just as effective research programmes or healthcare depend increasingly on teamwork, today we should all be business process analysts if we are to understand how competence in information adds value.

## A legal taxonomy

### Law firms must tackle a complex combination of challenges in their information management

Governance, compliance and chains of command, not to mention multiple languages, all add to the information management challenge for companies such as legal services firm Linklaters.

Information is a complex business at law firms because of the need for client confidentiality. Firewalls are important – content is organised around pieces of work for clients but these may involve different departments of the company. Companies may also have a global presence, as may their clients, and multilingual search is a big challenge.

So is it technology, or human and process issues that cause the biggest headaches? According to Liz Scott-Wilson, information architect at Linklaters, understanding governance and processes are the key challenges.

"Who's responsible for content creation and who for indexing and finding?" she says. "Where does escalation go? What are the interrelationships between information, content, users and repositories? How does information need to hang together, even if it is internally created? To help solve these issues, I am bringing in taxonomy management."

A taxonomy is about finding content: it's a hierarchical classification scheme that maps equivalent terms and the relationships between terms and concepts, explains Scott-Wilson. "It's essential in legal databases and for law firms – and in a multilingual law firm you need to map particular legal terms in more than one language and character set – which is a big challenge. Without a taxonomy, a search engine, however powerful, will not retrieve information usefully."

Search is still a problem for law firms, too, she says, no matter how big they are. "We use Autonomy for legal knowledge search. But I don't think any of the large practices have got it right yet, as we all have the same issues in trying to get a handle on internally generated knowledge.

"Externally provided information is quite well organised, by providers such as Westlaw and LexisNexis. We need to do a similarly good job on our own information."