Information Management

Martin White, Information Focus Ltd

Big data? – big opportunities!

Although there is a general awareness of the 'information explosion', there has been little analysis of just how rapidly business and society is generating information. Early In 2011 Hilbert and Lopez published a very detailed analysis of the growth of information storage, computation capacity and telecommunications capacity over the period from 1986 to 2007, based on over 1120 sources. Their analysis (The World's Technological Capacity to Process Information[1]) indicates that the storage of information has experienced a growth rate of 23% per capita over two decades. The main characteristic of the storage trajectory is the digitalization of previously analogue information, which has increased from 0.8% digital in 1986 to 94% in 2007.

All organisations track the extent of their assets, including their financial situation, the number of employees, equipment used in offices and factories, intellectual assets (patents, trademarks and licenses) and orders and deliveries. This information is required for regulatory reasons and to enable the Board of Directors to manage the organisation effectively. However, very few organisations have any measure of the amount of information that they have, even though this is a fundamental asset of the organisation in the 21st Century. Recently, MarkLogic published a very interesting survey of the extent to which senior managers understood the scale of the information problems that were about to arrive on their desktops[2]. The really scary outcome of the survey

[1] The World's Technological Capacity to Process Information

http://www.martinhilbert.net/WorldInfoCapacity.html

[2] The Post-Relational Reality Sets In. 2011 Survey on Unstructured Data

http://www.marklogic.com/news-and-events/press-releases/2011/marklogic-survey-reveals-

is that although 57% of managers regarded unstructured information as very important to the business, only 17% were highly aware of what information their organisation actually possessed. That is quite an information management gap.

One of the reasons why managers (even IT managers) have no idea of the amount of information stored in the organisation is that it is dispersed in so many individual repositories. In one of my clients recently I discovered they had 48500 Lotus Notes databases amounting to 45 terabytes of information. This represents something like 150 million pages, and yet the company (with a world reputation for quality) had no enterprise search application. Virtually all the information was invisible to any employee that did not, from personal experience, know which Notes databases to look at.

However 45 terabytes is nothing. The McKinsey Global Institute recently published a report entitled Big Data: The Next Frontier for Innovation, Competition and Productivity[1]. It runs to 156 pages, and every information professional should

unstructured-information-is-growing-rapidly-will-soon-surpass-relational-data.html

[1] Big Data: The Next Frontier for Innovation, Competition and Productivity

http://www.mckinsey.com/mgi/publications/big_data/index.asp

read it. According to McKinsey, analyzing large data sets—so-called Big Data—will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus. The authors of the report estimated that, by 2009, nearly all sectors in the US economy had at least an average of 200 terabytes of stored data per company (for companies with more than 1,000 employees), and that many sectors had more than 1 petabyte in mean stored data per company.

The problems of managing these volumes of information are not just technical. There is a major skills gap as well. McKinsey estimated that the supply in the United States of deep analytical talent in 2008 was around 150,000 positions. When they took into account current trends in new graduates with what they defined as deep analytical training (e.g., people taking graduate courses in statistics or machine learning, a subspecialty of computer science) this total rises to about 300,000. However, in a big data world, we expect demand for deep analytical talent could reach 440,000 to 490,000 positions in 2018—that's a talent gap in this category alone of 140,000 to 190,000 positions. In short, the United States will need an additional supply of this class of talent of 50 to 60 percent. Developing deep analytical skills requires an intrinsic aptitude in mathematics for starters, and then takes years of training. Addressing the talent shortage will not happen overnight, and the need for deep analytical talent that has already begun can only intensify.

These problems and opportunities are not just in the corporate sector. This is what the report has to say about public administrations in the EU:

Across Europe's public sector, the big data levers we identified for administration can bring three main categories of quantifiable monetary value as long as the right policies and enablers are in place. These are cost savings from operational efficiency, a reduction in the cost of errors and fraud in benefit administration, and an increase in tax receipts by narrowing the tax gap. Altogether, the public sectors of Europe's 23 largest governments could potentially create €150 billion to €300 billion—and potentially more—in new value annually over ten years. This implies boosting annual productivity growth rates in the public sector by about 0.5 percentage points above those expected, if current trends hold.

For the last couple of months I've been developing a techno-economic forecast for the enterprise search market in the EU. One of the trends that started to emerge was that search technology companies and companies trying to get the best from an enterprise search application were having a problem finding people with the right set of skills. These skills included:

- Highly skilled in software applications and Web 2.0 technologies

- Cognizant of the role that information plays in society

- Dedicated to providing a link between people and the information resources they need

- Active in research that is pushing information in new directions

- Specialist in the ethics and theory of information dissemination

- Adaptable to the information needs of businesses, organizations, and institutions.

This list comes from the University of Milwaukee[1] and is probably a good description of UKeiG members if you drop out the research element.

When I started to look for universities that offered a BSc in Information Retrieval I could not find any. There are around 25 information retrieval research groups in the EU, of which ten are in the UK, but they are almost totally devoted to PhD research projects and some MSc courses. What a missed opportunity.

Big Data is a relative term; it describes the way in which current ways of managing data and information become incapable of scaling up to the increase in volume. So the problems could be coming to the door of a small organisation in the near future, just as much as to a multinational company. I am convinced that Big Data is a big opportunity for the information profession. We will not be the sole owners of this buzz word but we could play a very important role.

By the way, do you know even roughly how much information is stored in your organisation, and

---

[1] University of Milwaukee

http://www4.uwm.edu/sois/programs/undergraduate/bsist.cfm

how much that has grown over the last year? If you don't perhaps you should, and then do something about it.