Access to information: Challenges and opportunities for the records profession

Dr Julie McLeod, Professor in Records Management, Northumbria University, UK

Introduction

The ability to access information begins with system design, and the act of accessing it begins from the moment of its creation. In the digital space access is increasingly easy and yet at the same time increasingly complicated. From a records management perspective access presents a number of non-trivial challenges and a range of opportunities. This keynote paper aims to consider some of these and the extent to which records management principles, practice and professionals can address them, what new approaches and partnerships are needed. It begins by considering what we mean by access and by records, moves on to examine some of the key challenges access to information presents in the digital world from a records management perspective, and then considers the role of the records professional in this space. This leads to the need for new approaches and partnerships which are illustrated by two examples of research conducted in the iSchool at Northumbria University (UK) that are relevant to the two conference themes.

The concept of access

In the recently published international *Encyclopedia of Archival Science* the entry for access states "The concept of access concerns whether an individual has permission or privilege to view or use a record or group of records (ISO 15489-1:2001). Users can include people, technology, and business processes that need to use records for a given purpose" (Duranti and Franks, 2015)[1]. Note the critical point that users are more than just human beings.

But what is the purpose or purposes of access? In the current climate of openness - open government, open data and big data - one would be forgiven for thinking the sole or main purpose of access to information was for transparency and accountability. This is a 'retrospective' purpose, so to speak, focused on holding governments, organizations and/or individuals to account. It is 'retrospective' in a similar way that access to information is for the purpose of writing and communicating history. However, for a records manager the most important purpose of access is more active, more immediate. It relates to the use and/or re-

[1] ISO 15489-1 defines access as "the right, opportunity or means of finding, using or retrieving information" which is more expansive.

use of information for *doing business*, whatever that is (e.g. government, service delivery, research, manufacturing, and education). Access to information supports decision making, service delivery, planning, innovation etc.; it supports efficiency (doing things right), effectiveness (doing the right things) and economy (the appropriate use of resources). For the records manager access is *not* primarily about compliance, regulation, or history, or at least they should not be the main drivers. If an organisation or individual does the right things, in the right way(s) and the data/information/records are managed appropriately then they will be there for transparency, accountability, compliance and historical purposes. This puts access to information firmly in the domain of the records manager rather than the archivist.

Dimensions of access

There are a number of dimensions to access for which information objects are at the core. The first is availability. Is the information discoverable and retrievable? Can we locate it and retrieve it through metadata and search tools? Is it available only in response to a request or is it proactively disclosed, for example as open data or through a Freedom of Information publication scheme as in the UK, and accessible through search or browse. Assuming it is available, is it also useable? Usability relates to interpretation, understanding and presentation. The third dimension of access is preservation, the implications of which are not solely the domain of the archivist interested in the long term retention of records, but also of the records manager for the shorter term retention and continued access to their information content. Despite the earlier fears of a digital dark age no longer being a concern according to some because it is tractable (Kilbride, 2011; Milic-Frayling, 2014), preservation is an important dimension if information is to be usable.

Access to information does not mean access by all to all, and hence rights of access is another dimension. This encompasses ownership, regulations, legislation, organizational and/or an individual's requirements. Access rights are challenging not least because they often change over time. For example, information about a planned merger or acquisition will be confidential and known only to a limited number of individuals prior to the merger; indeed 'Chinese walls' operate within financial institutions separating the investment function from the mergers and acquisitions function, to avoid any conflict of interest. Once the merger or acquisition is complete rights of access to the information/records will be less restrictive. Rights of access are managed through permissions and controls, mechanisms and processes that enable users (people, technology and business processes) to exercise their access rights

and to do so over time. This can be complicated. Related to this permissions and controls dimension is another - protection and security – how we protect the records/information through, for example, security classification and system controls. The international information security standard (ISO/IEC 27001, 2013) offers a best practice approach, at the centre of which is the preservation of the confidentiality, integrity and availability of information through risk management.

Crossing many of these dimensions is the issue of trust. Trust in what and whom; trust in the quality of the information/data and in the provider giving all of the relevant information; trust that the information will only be used in ways that were agreed or consented to, which is particularly important in the context of personal data and research data.

Characteristics of information

If these dimensions are to be supported, and the information is to be trusted, then information, or the 'information object', must display some important characteristics. It must be *authentic*, i.e. proven to be what it claims to be; to have been created or sent by the person claiming authorship and to have been created and/or sent at the time stated. It must have *integrity* to confirm it is a complete record that has not been altered or, if it has, that it is clear how it has been altered. It must be *reliable*, i.e. the contents can be confirmed as dependable, full and an accurate representation of the activity. And finally, as discussed earlier, the information should be *usable* i.e. as well as being locatable and accessible it can be understood and utilised through time.

Records

These characteristics are in fact the characteristics of records, defined in the international records management standard as "information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business" (ISO 15489-1, 2001, p3). Records are created as a result of some business activity or obligation and are kept as evidence of that activity as well as information to support it. ISO 30300 (2011, p8), the related standard for management systems for records, defines records as an asset, rather than information, and notes that evidence is "documentation of a transaction" and therefore "proof of a business transaction which can be shown to have been created in the normal course of business activity and which is inviolate and complete. It is not limited to the legal sense of the term."

Records have many purposes including supporting and documenting organizational policy formation and managerial decision making; providing consistent, continuous and productive management and administration; providing continuity in the event of a disaster; maintaining corporate, personal and/or collective memory and providing evidence of business, personal and/or cultural activity (ISO 15489-1, 2001, p.4). Records are a special form of recorded information that, when well-managed, can serve as instruments of accountability and authoritative sources of information for decision-making, planning, development, service delivery, rights management etc. Records tell the story.

However, to serve these purposes they need to have the four characteristics identified above i.e. they need to be 'good', quality records. It is this definition and these characteristics that are used to distinguish 'records' from 'information'. This distinction is well understood by records professionals, although personally I am concerned the definition is unnecessarily narrowing the scope of what records managers see as their domain and, in the world of e-discovery, lawyers have little concern about our terminology, only that evidence is found. I only note this here as Geoffrey Yeo explores the relationship between records and information in detail in his paper.

In summary, the scope of access is more than just 'getting at stuff', be that through search or proactive disclosure, more than accountability and information rights regulations. It is about supporting the business. Accountability, transparency and wider access should be a natural outcome. Access has a systems dimension (metadata, search tools, processes etc.), a human dimension (privacy, openness, sharing), a legal dimension (rights, laws and regulations across multiple jurisdictions and sectors) and an organisational dimension (supporting business, information culture etc.).

Challenges

In the digital space access has become increasingly complicated and presents a number of non-trivial records management challenges. For example, managing access to the unprecedented volume of information being captured; the unanticipated consequences of search and search engines on discovery; ensuring sufficient context in order to understand the information being accessed; balancing privacy, confidentiality and security with access, sharing and re-use in the world of open and big data; and defining the roles, responsibilities

and behaviour of information/records creators and consumers. Let us consider each one of these access issues and the extent to which records management principles, practice and professionals can address them.

First, *volume,* the digital iceberg. Figures suggest the information/data we create and copy is doubling in size every two years and will reach 44 zettabytes by 2020[2]. This digital iceberg is the result of changes in working practices in the digital world with global business, communication 24/7, less phone calls and face-to-face conversations, lack of thought and ease of cc/bcc and thank you emails; the result of digitisation for desktop or remote, mobile access, and of cheap storage. Virtual space has not suffered the constraints that physical space has, or the cost constraints as demonstrated by Kryder's Law[3], however the environmental and sustainability agenda may start to change this. It is an iceberg that is the result of easy retrieval, increasingly powerful software, and increased demand from users who are building new analytical tools to better understand consumers and customers, science, nature or society, for economic development (e.g. big data) or for intelligence (e.g. security, safety, espionage).

A large proportion of this information will be duplicates, i.e. redundant, and it will not be of equal value; much of it will have a transitory value. However, some of that volume is or can be a valuable asset; for example the volumes of environmental and climate records that show trends in weather patterns and help to predict weather events, or population data that helps us plan how to feed people; health records that enable us to track and fight disease etc. The tip of the iceberg may be what we see and can access, with the rest hidden, or perhaps the tip is what is valuable and the rest is not. From an organisational perspective it is not necessary to retain all information that is create forever, even for long periods, despite the fact that it seems easy to do so in the digital world. Nor is it necessary from a personal perspective and, I would argue, from a societal perspective. The vast majority of information / records, typically well over 90%, is not archived for permanent retention.

[2] 44 zettabytes is 44 trillion gigabytes. See: IDC. The digital universe of opportunities: rich data and the Increasing value of the Internet of Things. April 2014. http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

[3] Kryder's Law states that disc capacity doubles every two years and has translated into an exponential decrease in cost of digital storage over the past three decades. See: Walter, C. (2005). Kryder's Law. *Scientific American*, August, 293, p.32-33

In his book '*Delete: The Virtue of Forgetting in the Digital Age'* Viktor Mayer-Schönberger (2009) explores the phenomenon of "perfect remembering in the digital age," and why we "must reintroduce our capacity to forget... [our] ability and privilege of forgetting" (flyleaf). He opens with the story of Stacy Snyder who was not awarded her teacher's certificate because of a photograph she had posted on her web pages of herself at a party that was considered to be unbecoming of a teacher. There are many other examples of humiliating or damning content on social media sites that current or potential employers find and use with sometimes unwanted consequences, such as dismissal or failure to be appointed. Mayer-Schönberger (2009) "traces the important role that forgetting has played throughout human history, from the ability to make sound decisions unencumbered by the past to the possibility of second chances" (flyleaf) juxtaposed with "the monumental shift we are experiencing in the digital age, from a default of forgetting to one of remembering" (p13). What is his proposed solution? A records management tool! Expiration dates on information which he variously describes as "a modest response" (p189) and "crude" (p193). However, as records professionals know and, in fairness, Mayer-Schönberger acknowledges, delete is not straightforward.

From a records perspective the issues are recognizing what is important, ensuring the keep/destroy decisions are sanctioned and legal, and ensuring records are available and interpretable only for as long as required. This demands identifying what records should be created, kept and for how long, i.e. separating the wheat from the chaff and deleting what is no longer useful (some might say deleting what is useless). Decisions must be based on an assessment of the regulatory environment, business and accountability requirements and risks in order to meet organisatonal needs, compliance, and current/future needs of internal and external stakeholders. It requires retention decisions that meet regulatory requirements and are captured in a retention schedule that is then approved. It requires that records are disposed of according to those decisions, noting any litigation (legal hold) situations. It also requires a preservation strategy, which may need to address encryption for security and protection.

All of this concerns appraisal - determining the value of records (their information content) in order to make the decisions - and effective retention management. As Reed (2014, p. 127) wrote "access equals appraisal". In the digital world appraisal requires a risk managed approach which may be in the form of bigger buckets (less granular) or a selective approach such as NARA's Capstone approach to selecting the emails of certain roles/people for

retention[4]. Ideally the decisions need to be embedded at system design through metadata capture, but there are also opportunities for technology assistance e.g. forensics to de-duplicate; automatic classification of genres; AT&T's self-destructing email patents which seems to be based on the implementation of retention decisions and rely on good retention decisions[5].

This leads to the challenge of the *unanticipated consequences of being able to search* such volumes of digital information and records. From an access perspective, it is so much easier to search and find digital information and to make links and connections that would have been highly unlikely, if not practically impossible, in the analogue world. The consequences can be positive or negative. One positive consequence is access to existing information and data from many more sources. In a research context open data supports scrutiny and wider peer review as well as avoiding reinventing the wheel. It also enables different information objects and data sources to be linked more easily – digital with digital, digital with physical. For example, weather and/or environmental information with information about flora and fauna, or environmental and social information with health or disease, supports a greater understanding of the impact of climate on habitats and the effect of the environment on health. It enables researchers, governments, agencies and others to address the issues and to plan according to trends. In the big data context, access to data is predicted to support economic growth and the Internet of Things opens up a new area for records professionals. Another positive consequence is the development of alternative, more sophisticated search methods, such as semantic search and visual search, providing new access possibilities.

However, amongst the negative consequences is the identification of sensitive information through full text search that might not have occurred in the analogue world, other than perhaps serendipitously. In the online world personally, commercially or nationally sensitive information is easier to find; links are potentially easier to make. Last year saw a landmark ruling about Google and the discovery of personal information about a European citizen. The citizen complained "that an auction notice of his repossessed home on Google's search results, originally published in a Spanish newspaper in 1998, infringed his privacy rights

[4] See NARA Bulletin 2013-02. Guidance on a new approach to managing email records. 29 Aug 2013 http://www.archives.gov/records-mgmt/bulletins/2013/2013-02.html

[5] US Patent 8,725,809 13/052014. Method, system and apparatus for providing self-destructing electronic mail messages. Continuation of US patents 8,364,764 and 7,356,564.

because the proceedings concerning him had been fully resolved for a number of years and hence the reference to these was entirely irrelevant." He requested that the newspaper remove or alter the pages so his personal data no longer appeared and that Google remove his personal data so it no longer appeared in search results. The EU Court of Justice response required Google to "delete access to the information deemed irrelevant by the Spanish citizen, it did not rule that the content of the underlying newspaper archive had to be changed in the name of data protection (paragraph 88 of the Court's ruling). The Spanish citizens' data may still be accessible but is no longer ubiquitous. This is enough for the citizen's privacy to be respected" (European Commission, 2014). The notion of this 'right to be forgotten' is proposed in the draft new EU data protection legislation (European Commission, 2012). In Australia, the National Library's Trove online service[6] includes the ability to full text search digitised Australian newspapers, the contents of which are also already in the public domain. One of the site's FAQs is 'could you remove an article containing personal or family information?' The NLA's response is "We appreciate that some people are finding surprising information about themselves or their relatives which is sometimes good and sometimes bad, and that this may be of concern" but their disclaimer clarifies that they do not review or censor the newspaper articles[7].

From a records perspective search is concerned with ensuring that records have appropriate retrieval points and that the metadata about records is accessible over time; providing tools to assist in identification and retrieval of records; and ensuring appropriate access rights and accessible content. We must therefore determine what metadata should be created with the record and through the records processes, how that metadata will be persistently linked and managed. We must decide how to organise records to support users' browsing and search requirements and determine requirements for the retrieval, use and transmission of records between business processes and users. As the examples (above) illustrate it means we must identify potential sensitivities through a review and redaction process, and then validate or authenticate a person's rights to access a record, or a part of it, in relation to their role at a point in time. It requires mediation between the requirements of the person requesting access and the regulatory requirements and organisational rules, the implementation of appropriate controls on use(s) of the information, once permission to access has been granted. These may

[6] Trove http://trove.nla.gov.au/
[7] National Library of Australia. Trove FAQs http://trove.nla.gov.au/general/using-digitised-newspapers-faq/ and disclaimer "content which was published legally is not censored" http://trove.nla.gov.au/general/about#disclaimer

change over time due to the time sensitivity of the information contained in the record, or changing roles. Such requirements demand processes for managing security, dissemination and rights.

Ensuring *sufficient context* in order to understand the information being accessed is very challenging in the digital environment. From an access perspective, knowing where information came from, in what circumstances it was created, and its relationship to other information are vital to fully understanding it and establishing its authority. However, this is not always the case as two examples in science attest. Sir Cyril Burt was accused of falsifying data in reporting research on the heritability of intelligence as measured in IQ tests with twins Dr Andrew Wakefield's paper reporting research purporting to show a link between the MMR vaccine and bowel disease and autism, which caused many parents to refuse to have their children vaccinated and a rise in incidence of measles, was later shown to be "an elaborate fraud", grossly over interpreting the data. The paper was withdrawn from publication[8]. Whilst these go beyond context to publishing the data in full for peer review they demonstrate the importance of establishing authority through evidence and transparency

From a records perspective "[t]he context of records includes information about the business processes in which they are created. These metadata will allow users to understand the reliability of the record-creating authority [organisation/persons], the environment in which records were created, the purpose or business activity being undertaken and their relationships with other records or aggregations. But record metadata is not sufficient; records are managed in systems that are managed by organisations, which themselves operate in a broader context be it a business sector, government, nation, or a society). Sufficient information about these different layers is needed to make the records understandable and therefore useable to users (ISO 23081-1) and, since context may change, this information will accrue through time.

In the analogue world this is often apparent by looking at the record, or the file of which it is a part. Its form might signal a formal letter rather than a personal communication; the letter head would give the details of the organisation and possibly the department, the file an

[8] See for example: "The Burt Affair" (Sir Cyril Burt,1883-1971), https://en.wikipedia.org/wiki/Cyril_Burt; Dr Andrew Wakefield: Deer, B. (2011). Pathology reports solve "new bowel disease" riddle *BMJ*, 343 doi: http://dx.doi.org/10.1136/bmj.d6823

indication of the business process or activity. In the digital world this is often not clear. Email is the classic example with often missing attributions or subject lines, an incomplete email trail or multiple part-trails, and rarely any indication of the business process or activity to which it relates. Moss (2012; forthcoming) explores this very well.

Contextual metadata is very challenging because its creation and capture can be time consuming, if not expedited automatically through careful systems specification and design. Where the latter is not the case then there can be a burden on the creator who may not have adequate awareness or knowledge to effect its capture.

Balancing privacy, confidentiality and security with access, sharing and re-use in the world of open and big data is a complex and emotive area with tensions between the right to privacy, confidentiality and security (data protection) and the right to information. Facets include the need for governments to maintain security and the fears of a 'surveillance society'; the desire for governments to share information for innovation, economic and social benefit, and citizen engagement; the desire for the private sector to access data to improve services, target advertising etc.; the desire for the research community to stand on the shoulders of others and of the desire of funding bodies to encourage re-use and increase value of their investment. Sometimes these conflict.

Given the conference audience there is an interesting education example in the 2014 report on '*Big data and privacy: a technological perspective*' from the US President's Council of Advisors on Science and Technology. Access to the log information of online courses, including MOOCs, will make it possible to create and maintain longitudinal data about learner engagement with learning materials and activities, whether they repeat or skip content, their attention span etc., which when linked to grades will help improve education. But, if these are tracked over time and linked to an individual's future success then there are significant privacy issues: "[k]nowledge of early performance can create implicit biases that color later instruction and counseling. There is great potential for misuse, ostensibly for the social good, in the massive ability to direct students into high- or low-potential tracks (US Executive Office of the President, 2014 p14). As Richards and King (2014, p393) note "privacy protections focused on personally identifying information are not enough when secondary uses of big data can reverse engineer past, present and even future breaches of privacy, confidentiality and identity."

Two weeks ago the report '*A Question of Trust: Report of the Investigatory Powers Review'* was published in the UK (Great Britain. Independent Reviewer of Terrorism Legislation, 2015). At the behest of the British Prime Minister David Cameron, David Anderson Q.C. reviewed the effectiveness of existing legislation relating to investigatory powers, and examined the case for a new or amending law in the context of the threats to the UK, safeguards to protect privacy, challenges of changing technologies, and issues relating to transparency and oversight. The scope of his review was wider than counter-terrorism, considering the interception of communications, and collecting information about communications, missing persons investigations and crime for instance. Anderson made it clear that new laws are needed to cover security services' powers to monitor online activity, saying the UK needed "comprehensive and comprehensible" intrusive power rules rather than the existing multifarious "fragmented" and "obscure" legislation (Great Britain, 2015, p.4). Although Ministers want new laws to help police and agencies monitor online threats some have dubbed government proposals as a "snoopers' charter", warning the plans will infringe privacy. As part of an advisory group for a research network on information sharing in social care[9] I am acutely conscious of its potential benefits, but the ethics are challenging and, as with the 'snoopers' charter, there are issues of trust. Trust emerged as the key issue in an analysis of the discourse around the UK's *care.data* programme to collect and link together data from all health and social care settings (Childs and McLeod, 2015).

From a records perspective key issues are compliance with legal and regulatory requirements, applicable standards and organizational policy in this area; proactive and appropriate information sharing to reduce the number of information requests; and information security management. The requirements are some of those highlighted earlier viz. validating a person's right to access; managing the 'sensitivity' and access changes over time; designing and implementing permissions and security controls, and mediating between the requirements of the requester and the owner. All of these need to be agreed in policy. An interesting approach to addressing this from an archival perspective is the New Zealand Aotearoa Knowledge Creative Commons and Local Contexts work to support responsible navigation of archival records collections (Creative Commons, 2013; Local Contexts, 2014).

[9] Newcastle University ESRC Seminars Series (2014-17): Information Sharing in Policy and Practice: What needs to be shared (and not shared) when we share information? http://www.ncl.ac.uk/kite/esrc_seminars/

Balancing privacy, access, sharing and re-use is particularly challenging because it involves not only regulatory issues but also ethical and emotive ones. It is therefore not the sole responsibility of the records professional but is part of the wider information governance space inhabited by senior executives, lawyers, auditors, technologists etc.

The final challenge to highlight concerns the definition of *roles and responsibilities*. This is ultimately all about people, in particular understanding behaviour of information and records creators and consumers. People exhibit different behaviours and preferences; some are 'wired' to communicate using the latest technology to do so. For them tweets, texts and new forms of communication are preferred over email. For instance, a friend's children both use Facebook to arrange social meetings, but one uses it for planning them, the other to organise 'on the move'. This so-called Google generation may view records in different ways, if they view their communications as records at all. In part this leads to changing information culture(s), defined by Oliver and Foscarini (2014, p.11) as "the values accorded to information, and attitudes towards it." Add to this the impact of significant changes in organisational structures (flatter, open, more team based) on roles, culture, and behaviour which in turn are having an impact on perceptions about 'access'. There is a greater sense of a right to more information, certainly access to more information; citizens' expectations of their government and others and their ability to provide high quality information are greater. Together with national and social information cultures, these influence our understanding and discharge of our roles and responsibilities, and our capacity to discharge them effectively depends on our 'digital literacy'.

Digital literacy has been described as 'those capabilities which fit an individual for living, learning and working in a digital society' (JISC 2014). In an organisational context it encompasses the knowledge and skills (capability) of managers and staff to deal with information issues; their ability to use the technology *and* to understand the implications; and consequences of its use; their discernment i.e. their choices and trust in sources they access and their informed decisions about the information they share – how, why, when and with whom. Discernment in the digital age is not always in evidence or is sometimes partial. Proferes (2014, p.76) notes that some Twitter users seemed surprised and frustrated by "the seemingly newfound permanence of tweets" following The Library of Congress' deal with Twitter in 2010 to archive all tweets, making them available for anyone to read, embarrassing

content included. But, of course "tweets have never been fleeting" (p.77). It would be interesting to know how many Facebook users have opted to add a Legacy Contact, a feature announced in February 2015 which gives people a platform for remembering and celebrating the lives of loved ones when they die (Facebook, 2015), or have informed Facebook they would prefer to have their account permanently deleted after death.

If we suppose that it will be a records professional who will make access to information and good information and records management happen, then what key attributes and abilities should they reflect if they are to be effective in the dynamic digital environment? First they need to be an innovator and a risk taker, re-thinking the application of principles in practice, re-engineering services and systems, utilising technology in effective and imaginative ways, and taking a risk-assessed approach. We cannot and do not need to apply the highest 'gold' standard to the management of *all* records; managed risks are required. Records professionals must demonstrate leadership in the access to information arena, else others will (e.g. IT and information government professionals), and they should be collaborators. Partnering with computer scientists, mathematicians, lawyers and psychologists, for instance, is essential because of the complexity and multi-disciplinary nature of the digital challenges. Access to information is a part of the broader information governance and technology landscape of which records management is one facet. Finally, today's records professional needs to be an expert communicator and educator, communicating the value of records management in this space and educating the consumers and creators (the new records managers) so that they are digitally literate and have the necessary knowledge and skills to discharge their information responsibilities. Records managers no longer manage records for others but facilitate their management through systems and processes using a range of standards, tools and metrics. However, in light of the issues discussed here records managers need a much better understanding of information behaviour and digital literacy if they are to facilitate the education of others. We can learn a great deal from the decades of research and expertise of librarians this domain. These qualities are generic to any profession; records professionals must reflect them.

The value of managing records

"In most instances contemporary records management is conducted within organizations devoid of any connection or consideration for archival concerns" (Duranti and Franks, 2015). Rather, as stated earlier, it is as much about supporting current business functions and

processes as supporting the ability of the organization to respond to its accountability requirements. Records professionals must, therefore, position themselves to maximise their support for the achievement of organisational goals and objectives. This requires a thorough understanding of the organisation and the context in which it operates (drivers etc.); it means identifying priorities, risks and requirements and articulating the value proposition, which may change over time. For example, in the pharmaceutical sector records are managed for compliance with regulations to enable a new drug to enter the market (e.g. Food and Drug Administration). However, reducing the time to market through more effective internal access to information, and hence increasing sales and profit before patent expiry and generic drugs are developed, is the 'value proposition' for records management. Envisioning this and communicating it requires the qualities mentioned above.

Different approaches and partnerships

Some of our existing, traditional records management principles are adequate and appropriate for addressing the challenges, for example appraisal and retention, but the way they are implemented needs to be reinvented because of the characteristics of the digital space - speed, dynamics, value etc. Two examples of research and development work at Northumbria University to tackle some of these challenges highlight a different, innovative approach and different partnerships. The first, technology assisted sensitivity review of records, relates to the conference theme of records management and its connection to society; the second, research data management, relates the theme of records management and research, though both examples have some relevance to both conference themes.

Technology assisted sensitivity review of records

In the UK the Public Record Act (PRA) assumes that records will eventually be publicly accessible and, on transfer to The National Archives (TNA), will be open unless there is any reason for closure of part(s) of them. Government departments must review the records that have been selected for transfer against public criteria contained in, for instance, TNA policy, the PRA, and the Freedom of Information (FoIA) and Data Protection Acts, and against FoIA exemptions for enduring sensitivity. 2017 will see the first significant volume of digital information as a result of the change in the 30 year rule for transfer to 20 years, and there are some significant issues. It is anticipated that the volume of deposits will increase, possibly four-fold from 5% to 20% of records captured. Changes in work practices have led to fragmented information, for example in email threads as discussed by Moss (2012), making

review more difficult and its cost greater. In the context of search making it easier to discover and make connections (some of which could be sensitive) in the digital world, if appropriate and accurate sensitivity review cannot be assured there is a risk this will lead to precautionary closure. There are implications for social and historical research and the (potentially more limited) ability of citizens to challenge conclusions and hold government to account. To avoid such a situation there is a need to reduce the cost of digital records review and to increase throughput whilst maintaining or improving the quality of the review process.

One approach is to use sophisticated information retrieval algorithms that employ techniques such as archival diplomatics to identify potentially sensitive information, by looking for names that might be sensitive or combinations of entities that could identify individuals, such as a name and date of birth, a role, a place etc. (Moss, forthcoming). Project Abacá, a feasibility project between Glasgow and Northumbria universities, has explored technically assisted sensitivity review of UK digital public records by developing such algorithms. Still nascent, these and others under development elsewhere, "will be able to distinguish sensitive information at only the most simplistic level, such as an insurance number or details of a bank account; all other instances that are flagged will need to be reviewed. They will be able to rank sensitivity, prioritizing instances of possibly the highest sensitivity" (Moss, forthcoming). However, they have potential and they illustrate an innovative approach through a different collaboration between records professionals and computer scientists. They also have potential application in other contexts and sectors, such as e-discovery.

Research data management

Open research data is part of the open data movement and is required by UK and other national research funding bodies. Research data is made open for two main reasons: to provide evidence that the research was conducted properly (witness the earlier science examples), and to provide data for reuse (secondary analysis), generating further findings and outputs i.e. 'standing on the shoulders of giants'. Research data can be made available for these two purposes without being open of course, for instance via controlled access for designated people only.

Recognising the provision for research data management (RDM) in the UK varied between different disciplines and there was a "shortfall" in technical and human infrastructure, in 2009 JISC (Joint Information Systems Committee) announced the first of two multi-million pound

Managing Research Data programmes (JISC, 2013). Between 2009 and 2013 these programmes funded projects to identify the requirements for, and then build, infrastructure for effective RDM in UK universities. Northumbria University's iSchool undertook two projects. The first, '*DATUM for Health*', developed a research data management skills training programme for postgraduate research students in health studies; and the second, '*DATUM in Action*', supported researchers to plan and implement RDM in practice (Northumbria University, 2012). Both projects involved new partnerships. Partners in the first one were the university's Graduate School (responsible for PhD student training and development), staff and students in the School of Health, and the Digital Curation Centre (DCC) and Digital Preservation Coalition (DPC), two leading UK digital curation/ preservation bodies. Partners in the second one were academic colleagues in health, mathematics and computer science. Advisory boards for each project included other stakeholders such as the university's central research services and the library.

The project aims were, respectively, to enhance the knowledge and skills of PhD students (new researchers) in managing their research data and to improve RDM in practice. RDM involves a lot of records management but goes beyond that, for example covering concepts of anonymisation and consent. However, the approach taken was to focus on the researcher and their research process, rather than on records management. By making it clear that managing data was as much part of the research process as (say) methodology, that it was the proactive, planned management of research data throughout the research process, from proposal to publication, and that it supported the researcher in doing their research, it became meaningful and more valued. The key tool to achieving RDM is the 'data management plan', a structured document or tool in which to record decisions and details about how research data will be created and captured, managed, shared, protected and preserved. Feedback from the training developed for the PhD students suggested RDM was not something with which they were familiar and that in considering what data would be created and how it would be used, more informed decisions were made about how it should be collected and what consent might be required for potential further use. Insights arising from the researchers on the second project outlined particular RDM issues for qualitative data with a health focus in the open data context. These non-trivial issues related to methodology, ethics and practicalities and provide new opportunities for records professionals in a research context (Childs et al., 2014). In addition to advising and providing more guidance on data/records appraisal, retention and destruction, records professionals can advise on the costs of preparing and curating data for

open access, which can outweigh its value, and on storage and access mechanisms, particularly where national or institutional repositories do not exist or do not yet hold research data; they can also advise on consent and anonymity and the implications for subsequent information sharing. Records professionals can collaborate with IT staff and librarians to develop the necessary supporting human, procedural and technology infrastructure and, perhaps most importantly, can demonstrate *leadership* as the DATUM projects did in the context of RDM at Northumbria University, by seeking partnerships to make it happen.

Conclusion

Access to information is not only complicated but also complex because of the human dimension. Records management principles, such as appraisal, retention management, metadata capture, permissions and security frameworks, can and do support access to information but there are some non-trivial challenges which I have only been able to scratch the surface of here. The challenges present opportunities which, as the examples from my own university illustrate, require new approaches, such as risk-based, proportionate ones, information/records creator focused ones and technology assisted methods, and new partnerships.

In a country where football is important to use a football analogy we need to take the ball and run otherwise we will be intercepted and left behind. I look forward to many more examples, views on these and other challenges, and opportunities over the course of the conference.

References

Childs, S. and McLeod, J. (2015). *A case example of public trust in online records: the UK care.data programme*. Interim report of a project undertaken for the InterPARES Trust project. https://interparestrust.org/

Childs, S., McLeod, J., Lomas, E. and Cook, G. (2014). Opening research data: issues and opportunities. *Records Management Journal*, 24 (2), p.142-162.

Creative Commons Aotearoa New Zealand (2013) *Indigenous knowledge.* http://creativecommons.org.nz/indigenous-knowledge/

Duranti, L. and Franks, P. (Eds), (2015). *Encyclopedia of Archival Science*. Rowman and Littlefield.

European Commission. (2012). *How does the data protection reform strengthen citizens' rights?* http://ec.europa.eu/justice/data-protection/document/review2012/factsheets/2_en.pdf

European Commission. (2014). *Factsheet on the "right to be forgotten ruling" (C131-12)*. http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

Facebook. (2015). *Adding a legacy contact*. Press Release 12 Feb 2015 http://newsroom.fb.com/news/2015/02/adding-a-legacy-contact/

Great Britain. Independent Reviewer of Terrorism Legislation. (2015). *A question of trust. Report of the investigatory powers review*. [The Anderson Report]. 11 June 2015, Stationery Office https://www.gov.uk/government/publications/a-question-of-trust-report-of-the-investigatory-powers-review

ISO 15489-1 (2001). *Information and documentation. Records management. Part 1: General.* ISO.

ISO/IEC 27001. (2013). *Information technology. Security techniques. Information security management systems. Requirements*. ISO.

ISO 30300 (2011). *Information and documentation. Management systems for records. Fundamentals and vocabulary*. ISO.

JISC (2014). *Developing digital literacies*. Available at: https://www.jisc.ac.uk/full-guide/developing-digital-literacies

JISC (2013). *Managing research data programme 2011-2013*. Available at: http://www.webarchive.org.uk/wayback/archive/20140614021511/http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx

Kilbride, W. (2011). Aiming for obsolescence. (Editorial). *What's New*, April. Digital Preservation Coalition. http://www.dpconline.org/newsroom/whats-new/684-whats-new-issue-35-april-2011#Editorial35

Local Contexts (2014). *Traditional knowledge licences.* www.localcontexts.org/

Mayer-Schönberger, V. (2009). *Delete: The virtue of forgetting in the digital age*. Princeton University Press.

Milic-Frayling, N. (2014). Sustainable computation as a means for digital preservation. *2nd Annual Conference of the ICA, Girona, October 2014*. Pre-paper notes in: Evaluation and strategies of digital preservation and UNESCO's role in facing the technical challenges van Gorsel, M., Leenaars, M., Milic-Frayling, N. and Palm, J. http://www.girona.cat/web/ica2014/ponents/textos/id100.pdf

Moss, M. (forthcoming). What is the same and what is different. In: Moss, M, & Endicott-Popovsky, B. (Eds). *How information creation, capture, preservation and discovery are being transformed*. Facet Publishing.

Moss, M. (2012). Where have all the files gone? Lost in action points every one? *Journal of Contemporary History*, 47 (4). pp. 860-875. http://nrl.northumbria.ac.uk/13176/

Northumbria University. (2012). *DATUM: research data management*. Available at: http://www.northumbria.ac.uk/datum

Oliver, G. & Foscarini, F. (2014). *Records management and information culture: Tackling the people problem*. Facet Publishing.

Proferes, N. (2014). What happens to tweets? Descriptions of temporality in Twitter's organizational rhetoric. In: *iConference 2014 Proceedings* (p. 76–87). doi:10.9776/1404

Reed, B. (2014). Reinventing access. *Archives and Manuscripts*, 42 (2), p. 123-132.

Richards, N. and King, J. (2014). Big data ethics. *Wake Forest Law Review*, V49, p.393-432. http://pacscenter.stanford.edu/sites/all/files/RIchards%20and%20King%20Ethics.pdf

USA. Executive Office of the President. President's Council of Advisors on Science and Technology. (2014). *Big data and privacy: A technological perspective*. US Government. https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf