eLucidate

Vol. 15 Issues 1-2, Autumn 2018

ISSN: 1742-5921

eLucidate's editor is Gary Horrocks email: *info.ukeig@cilip.org.uk*

eLucidate is published by UKeiG, the UK einformation Group.

UKeiG is a Special Interest Group of CILIP: the library and information association. CILIP Registered Office: 7 Ridgmount Street, London, WC1E 7AE, Registered Charity No. 313014

Contents

Editor's Note

Welcome to Volume 15 issues 1-2 2018 of UKeiG's journal eLucidate.

A huge vote of thanks to the expert contributors who have made this special double issue of eLucidate so thought-provoking and informative. Many of the articles celebrate the ongoing success of our continuing professional development programme and highlight our upcoming courses for the remainder of 2018.

Karen Blakeman laments the dumbing down of the search functionality of Google. Are its days as a research tool over? She identifies the "useless and irrelevant" pursuit of revenue generating mobile personalisation options as the culprit and rally cries that now is the time to consider alternative search tools.

Andrew Cox reflects on the growing importance of data in the work of academic librarians and information professionals. He articulates a "data role spectrum" that embraces familiar skills sets like information/data literacy, training, collection management and metadata, but throws many unfamiliar competencies into the mix: data curation, integrity, analysis and visualisation. Data management offers huge opportunities to extend and enhance the skills and knowledge academic librarians have had for years. New wine in old bottles perhaps.

David Ball builds on this theme with an exposition of the emergence of Open Science and Open Data. Notably he emphasises that this developing paradigm goes way beyond STEM subjects and impacts on social science, arts and humanities. Research data can be qualitative and well as quantitative and embrace statistics, digital images, archives, sound recordings and survey data, for example. He provides an extensive overview of Open Science and Open Data, their rationale and potential. This article is a taster of his successful UKeiG CPD course – "Open Access, Open Data, Open Science: Anatomy of a disruptive technology" - which we hope to repeat early next year.

Martin White emphasises the relevance of academic research on the success of enterprise search projects. "Academic research may not provide definitive answers to very difficult issues but it can provide a vendor-independent framework for discussion and inspiration." He emphasises valuable research that has significant implications for enterprise search success, introducing us to concepts like "information scent" along the way. If you have responsibility for search management in your organisation, Martin's article is a must read.

Marlize Palmer O.B.E, UKeiG's Information Manager of the Year (2016) and James Dawes-Hughes explore the role a Library Management System can play in the archiving and preservation of tweets. Twitter has a huge role to play in research, linking information from a variety of sources including emails and Word documents.

We also have our regular online resources update and an insight into the rationale behind Internet Librarian International's upcoming twentieth birthday conference.

This issue concludes with a celebration of our prestigious Strix Award as we prepare for the 2018 Annual Memorial Lecture in London on Friday 23rd November 2018. We hope to see you there. Until then, enjoy this issue and please give us your feedback and join us in discussions on our Twitter, LinkedIn and Facebook accounts.

An archive of previous issues of eLucidate is available here. We endeavour to feature contributions from experts in the field, keeping members up to date with developments and innovations in the digital information industry, considering the impact on information professionals and consumers of e-information. Core topics for consideration include: digital literacy, effective information retrieval and search technologies, intranets, social media, open access, e-publishing and e-industry research and development.
UKeiG encourages the submission of articles and reports about any of the topics covered by the journal, and contributions and suggestions for content can be emailed to me at: *info.ukeig@cilip.org.uk*

Pease refer to Notes for Contributors for further information.

Editor – eLucidate
info.ukeig@cilip.org.uk
LinkedIn
Twitter

The future of Google is not about search

Karen Blakeman, RBA Information Services

karen.blakeman@rba.co.uk

Google's recent and seemingly innocent announcement that it is spring cleaning time for Blogger generated alarm amongst many of its users, including myself. The move is intended, we are told, to simplify the platform and *"to enhance the blogging experience for all of our users."* This could be the beginning of the end for the blogging service. We've seen it all before.

Look at Google News, now a shadow of its former self: no more advanced search, useless and irrelevant personalisation options, and don't even think about trying to set up sensible alerts. Alerts have never worked that well at the best of times but have not improved one iota since News was revamped. Google Finance has gone the same way: no more portfolios for monitoring stocks, no more historical data for viewing and download, no more news annotations on the price charts, and the comparison option only works for two stocks at a time. If you are interested in monitoring the stock markets or researching individual companies for free get thee hence to Yahoo! Finance. There was some doubt over the future of Yahoo! Finance when Yahoo! was acquired by Verizon and became part of Oath but, charting oddities aside, there does seem to be some development going on. The new "Sustainability" tab, for example, shows environment, social and governance (ESG) ratings from Sustainalytics.

One of the main drivers of change for Google and other search engines is Mobile-First. Google is very much at the forefront of prioritising mobile access to information and is bent on pushing what it thinks you need to know for your daily routine: routes to work, location of road works, train times, do you need an umbrella today? (Forget Facebook, Google hoards even more of your personal data – see BGR.) It is Mobile-First that is responsible for the loss of easy access to country versions of Google.

Google's search results are now based on your current location. What's new, you may ask? Google has always looked at location, even down to city/town level, and changed search results accordingly. That's fine if I am travelling and want, for example, to find the nearest Thai restaurant via my phone. Presenting me with a list of eateries in my hometown of Reading is no good if I'm working in Manchester and getting very hungry. The problems start if you are researching a person, company or industry based in a country other than your own – let's use Norway as an example – or just want the latest news from that country. The trick used to be to go to the relevant country version of Google, in this case http://www.google.no, run your search and Google would give preference to Norwegian content. It was, and still is, a great way to get alternative viewpoints on a

topic and more relevant local information. Now, regardless of which version of Google you go to, you will only see results tailored to your current physical location.

In a blog posting - "Making search results more local and relevant" - Google says:

"Today, we've updated the way we label country services on the mobile web, the Google app for iOS, and desktop Search and Maps. Now the choice of country service will no longer be indicated by domain. Instead, by default, you'll be served the country service that corresponds to your location. So if you live in Australia, you'll automatically receive the country service for Australia, but when you travel to New Zealand, your results will switch automatically to the country service for New Zealand. Upon return to Australia, you will seamlessly revert back to the Australian country service."

It confirms what many of us had long suspected: that mobile search is what Google is concentrating on. It is, after all, where most of Google's revenue comes from. There is a way around it but it is rather long-winded. You need to go to Google's Settings using either the link in the bottom right hand corner of the Google home page, or the one near the top of a search results page, and then click on Advanced Search. On the Advanced Search screen scroll down to "Then narrow your results by…" and use the pull down menu in the region box to select the country. You may additionally need to change the language.

Alternatively, you could use a VPN or the Tor browser to make it look as though you are in another country or include the *"site:"* command in your search. If we were looking for Norwegian based sources we could use *"site: no"* but that would, of course, miss Norwegian sites registered as .com or with other international domains. Phil Bradley has carried out a useful comparison of the various options in his posting "Google improves search; makes it much harder."

Google says:

"We're confident this change will improve your Search experience, automatically providing you with the most useful information based on your search query and other context, including location."

No, Google. You have just made things a lot more difficult for those of us who conduct serious, in-depth research.

There are search engines other than Google, and specialist tools and portals for specific types of information and subject areas. As Google's search services are continually degraded and functionality removed, ensuring that you have back up search strategies in place and becoming aware of resources that can take you directly to information in your subject area are becoming increasingly important. I recently facilitated a workshop where we were exploring not just Google but also alternative search tools and different

approaches to tackling search, including deep web and even the dark web. On the same day Google delivered a two-hour keynote presentation at its annual developer conference, in which Artificial Intelligence (AI) technologies were the focus and search was barely mentioned. (Google Assistant takes centre stage at I/O, search takes a back seat.)

The company's efforts seem to be concentrated on developing Google Assistant and introducing new features such as Duplex, which is an AI driven system for making "voice" telephone bookings and scheduling meetings. (Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone.)

"Today we announce Google Duplex, a new technology for conducting natural conversations to carry out 'real world' tasks over the phone. The technology is directed towards completing specific tasks, such as scheduling certain types of appointments. For such tasks, the system makes the conversational experience as natural as possible, allowing people to speak normally, like they would to another person."

Google obviously thinks that a quick phone call to the hairdresser to make an appointment - one of the examples demonstrated at the conference - is too onerous a task for mere mortals! And if you find writing emails a chore there is Gmail Smart Compose. (Write emails faster with Smart Compose in Gmail.)

"*From your greeting to your closing (and common phrases in between), Smart Compose suggests complete sentences in your emails so that you can draft them with ease ... It can even suggest relevant contextual phrases. For example, if it's Friday it may suggest 'Have a great weekend!' as a closing phrase.*"

What could possibly go wrong?

AI and machine learning are also being applied to News and Maps, and RankBrain has been used for a while as part of search to help sort and organise results. (See: A Complete Guide to the Google RankBrain Algorithm.) Despite this Google still gets it wrong, not because the technology has failed but because, we are told, people are confusing it! (Google says it is struggling to "understand truth" because people are confusing its search algorithm.)

All of this would be laughable if more sinister possibilities were not possible. A video produced for internal viewing within Google in late 2016 was recently leaked and imagines how data collection by Google could influence users into actions to help achieve their goals, or even guide the behaviour of entire populations to solve global problems. Entitled the "Selfish Ledger", it is scary stuff. (See the article "Google's Selfish Ledger is an unsettling vision of Silicon Valley social engineering", published on The Verge website.) According to a spokesperson from X (formerly Google X) it is intended to be disturbing. It is a thought experiment using a technique called speculative design to explore uncomfortable ideas to provoke discussion and debate. "It's not related to any current or future products". On the contrary, looking at recent announcements and presentations from Google, it seems that some of the concepts of the Selfish Ledger are beginning to appear in its products and services.

I urge you to read The Verge article and watch the video (viewable from within the article) and then look at what is happening to Google's services. Google's future is not about search. Perhaps now is a good time to consider alternatives?

Karen Blakeman's UKeiG CPD course Navigating the Deep Web: Advanced Search Strategies for Researchers is on Thursday, 15th November 2018 at CILIP headquarters in London.

Academic librarianship as a data profession: The familiar and unfamiliar in the data role spectrum

Andrew Cox, Senior Lecturer, Information School, University of Sheffield

a.m.cox@sheffield.ac.uk

Summary
The paper reflects on the growing importance of data in the work of academic librarians. Based on tasks related to research data management (RDM) it proposes a data role spectrum, ranging from familiar tasks such as supporting search, training and collection management, through to less familiar tasks, such as data analysis and visualisation. Areas where library work is increasingly tied up with data, such as Text and Data Mining and bibliometrics, can be analysed for their positions on the spectrum.

We are used to thinking of data as sitting beneath information and knowledge in a pyramid of value. But the concept of "data" has grown significantly in importance in the last five years, driven by public speculation about the power and risks of big data. In parallel, data related roles are becoming increasingly important in professional practice. Librarianship is becoming more of a "data profession". One obvious starting point for thinking about this impact in the academic library context is to analyse the specific case of Research Data Management (RDM). This is an area where dealing with data is clearly central. But is this familiar territory, just reinventing or extending what librarians already do or does it require a new set of competencies?

The most direct way to understand what Research Data Management means to academic librarians in practice is to think of some of the key tasks that are involved in delivering a research data service. If one were to make a shortlist it would certainly include: helping a researcher to find pre-existing data sources relevant to their research; running a training or awareness session; reviewing metadata associated with a potential deposit into a data repository; investigating what researchers need in terms of support; inputting to the creation of a data policy framework; and offering advice on a Data Management Plan (DMP) for a project proposal. Nearly all of these activities have strong continuities with what we already expect to do as a librarian.

What could be more natural for a librarian to find themselves helping someone search for and evaluate a source, albeit data rather than a published text? It is true there are a few differences from supporting the usual literature search tasks. One needs to know a bit about what data sources and data archives there are out there. There might well be licence conditions associated with data reuse that need careful analysis, in a way not applicable to published texts. Further, a source of data might not necessarily be ready packaged: it could be that there is an API or some data service through which the

researcher can generate a relevant dataset. However, overall this feels like a very familiar role that a librarian would have the requisite skills for (Gregory et al., 2018).

Adding something about RDM to an existing user training session is again the familiar ground of academic librarianship, and its information literacy focus. In all likelihood, the objectives of such a session would be to make researchers aware of institutional, funder and publisher requirements; to make the case for data sharing, albeit qualified by a sense of potential exceptions; and explore practical data management basics (such as back-ups and file-naming). So it would require acquiring new knowledge, yet there are strong continuities with open access advocacy, and learning outcomes around these topics would fit into existing information literacy sessions for researchers.

Reviewing a potential deposit to the data repository is again familiar territory. It is a fairly standard library task rooted in collection management principles and based on an understanding of the importance of metadata and standards. It might well be combined with the role of monitoring metadata related to outputs in the repository.

Gathering requirements from researchers about their support needs also should come naturally to a user service focussed profession like librarianship. Librarians are used to gathering data from interviews, focus groups and surveys to try and discover what services users need, and then designing services or procuring systems to meet these needs. RDM takes us deeper into the research process as an aspect of scholarly communication, but a strong interest in user behaviour is a good starting point for carrying through this task. Data policy is about creating a governance structure within which data is valued and managed. Contributing to the development of such a policy is probably a fairly familiar task, and requiring a good understanding of the wider policy context, such as relevant institutional and national policies.

Helping a researcher write their DMP could be the most unfamiliar of all the tasks I listed above. While the skills needed to do it effectively are the same as any advice service, this particular support requires a fairly deep knowledge of funder requirements; of relevant standards; of local data management processes (e.g. around data storage); as well as a feel for the research process.

Thus so much of what RDM is about could be considered somewhat familiar territory. It seems to involve acquiring new knowledge but much of the role is familiar. There are plenty of ways that roles in RDM build on skills and knowledge that most librarians already have (Cox and Verbaan, 2018). However, there are some more tasks that could potentially be involved in supporting RDM. If we look at the other end of the spectrum from the common tasks, we could also identify potential roles in data curation, data carpentry, data integrity, data analysis & visualisation and also embedded roles in research project teams. These are more like the specialist or cutting edge of RDS. They would include:

- Data curation is the long-term digital preservation of datasets. Preservation has traditionally been an aspect of library work, but it is perhaps more in the territory of archives.
- Data carpentry is about understanding how to manipulate and transform data, preparatory to analysis.
- Data integrity links to the traditional interest of information professionals in data quality, but in the context of a crisis of reproducibility (in certain subjects), the assurance perhaps lies with more open science.
- Embedded roles involve working directly with a research team. This is more about how the role is delivered than the knowledge/expertise required. It is about breaking out of the library and working with researchers on a daily basis.
- Supporting or undertaking data analysis and visualisation, or at least having a role in the selection and supporting use of computational tools to do analysis.

It remains to be seen whether academic libraries will start to see these roles as standard tasks. Probably institutions will vary depending on their research-intensive nature, among other factors. Emerging from this discussion is what I am calling the "data role spectrum". The data role spectrum starts with the familiar and transitions to the unfamiliar.

- Support for data search/access to data
- Data literacy training and promoting awareness
- Data collection management, including metadata
- Gathering support requirements for services/tools
- Data policy
- Data management advice
- Data carpentry
- Data curation
- Data integrity
- Embedded roles in a research team
- Data analysis and visualisation

At one end of the scale are the data related tasks that feel relatively familiar, such as data search, data literacy training and data collection management. At the other end of the scale are those that we might associate with the work of IT professionals, such as data carpentry, or researchers, such as data analysis and visualisation. These fit less easily into the classic library role, but the profession is changing to engage with them more. I think there could be very different perceptions of how "familiar" some of these roles are and there is a lot to unpack under the heading of something like data curation, but the spectrum is useful for looking at other areas where the "deluge of data" is flooding into academic library work. Here are a few examples. One emerging area is Text and Data Mining (TDM) where machine-learning tools are used to analyse huge unstructured corpuses of texts. In TDM the library role is most likely to be about licensing content (data collection management) and training people to use tools (data training), and in managing derived outputs (a form of data curation). Another area that could be analysed through the spectrum is around bibliometrics and altmetrics, where "the data" is about researchers and their outputs, analysed to reveal the impact of research. Here the role seems mostly to be about calculating metrics, and helping researchers to understand

metrics about themselves and their work, in the context of the concept of responsible use (Cox et al. 2017; Bibliomagician, 2017). Interestingly, here the role is actually in the area of doing analysis/visualisation, albeit, through proprietary tools, and with an emphasis on training others to undertake analyses for themselves. One could do a similar review for other areas such as library and learning analytics, although here it is about analysing data to support management decisions, rather than supporting users.

Academic librarianship seems to be moving towards becoming a data profession. It will be interesting to see how things develop over the next decade. Perhaps data analysis will become integral to professional competencies. This would be a fundamental shift in the positioning of academic librarianship. There are much more obvious areas where including data in support services or training is a natural extension of what academic libraries already do; and other areas such as data curation that others would expect librarians to take up. The data role spectrum could help chart these changes.

References

Bibliomagician (2017). *Competency model for bibliometric work*.

Cox, A.M. and Verbaan, E. (2018). *Exploring Research Data Management*. London: Facet.

Cox, A., Gadd, E., Petersohn, S., & Sbaffi, L. (2017). Competencies for bibliometrics. *Journal of Librarianship and Information Science*.

Dooley, J. M. (2015). *The archival advantage: Integrating archival expertise into management of born-digital library materials*. Dublin, OH: OCLC Research.

Gregory K, Khalsa SJ, Michener WK, Psomopoulos FE, de Waard A, Wu M (2018) Eleven quick tips for finding research data. *PLoS Computational Biology* 14(4)

Andrew Cox and Eddy Verbaan's UKeiG CPD course, Research Data Management for Information Professionals, is on Wednesday, 14th November 2018 at CILIP headquarters in London

The emergence of open data

David Ball, David Ball Consulting

davidball1611@gmail.com

Summary
Open Science is moving centre-stage, with a rationale of improving efficiency in science; increasing transparency and quality in the research validation process; speeding the transfer of knowledge; increasing knowledge spill-overs to the economy; addressing global challenges more effectively; and promoting citizens' engagement in science and research. Open Data has undergone a surge in practical development, mirroring the well-established repositories for research outputs. The development and application of model policies and of principles are discussed and the views of researchers championing Open Data highlighted.

1. Open Science

1.1. Why Open Science?

The concept of Open Access (OA) to research outputs such as journal articles has been common currency for many years. The seminal Budapest Open Access Initiative (BOAI) statement of February 2002, for instance, reads:

> *"By open access to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."*

More recent thinking, however, for instance by the European Commission, has expanded the concept of openness even further, to Open Science (OS), which aims to transform science by making research more open, global, collaborative, creative and closer to society. This shift is potentially very important for the development and exploitation of research.

OS is about the way research is carried out, disseminated, deployed and transformed by digital tools, networks and media. It relies on the combined effects of technological development and of cultural change in the direction of collaboration and openness in research.

To elaborate, an OECD report (1) identifies the following six rationales for policies that seek to implement and support OS, including Open Data:

- *Improving efficiency in science* – OS can increase the effectiveness and productivity of the research system, by: reducing duplication and the costs of creating, transferring and re-using data; enabling more research on the same data; multiplying opportunities for domestic and global participation in the research process.
- *Increasing transparency and quality* in the research validation process, by allowing greater replication and validation of scientific results.
- *Speeding the transfer of knowledge* – OS can reduce delays in the re-use of the results of scientific research, including articles and data sets, and promote swifter development from research to innovation.
- *Increasing knowledge spill overs to the economy* – Increased access to the results of publicly funded research can foster spill overs and boost innovation across the economy as well as increase awareness and conscious choices among consumers.
- *Addressing global challenges more effectively* – Global challenges require co-ordinated international actions. OS and Open Data can promote collaborative efforts and faster knowledge transfer for a better understanding of challenges such as climate change, and could help identify solutions.
- *Promoting citizens' engagement in science and research* – OS and Open Data initiatives may promote awareness and trust in science among citizens. In some cases, greater citizen engagement may lead to active participation in scientific experiments and data collection.

1.2 What it is

Each step of the research lifecycle is becoming more open, for instance through:

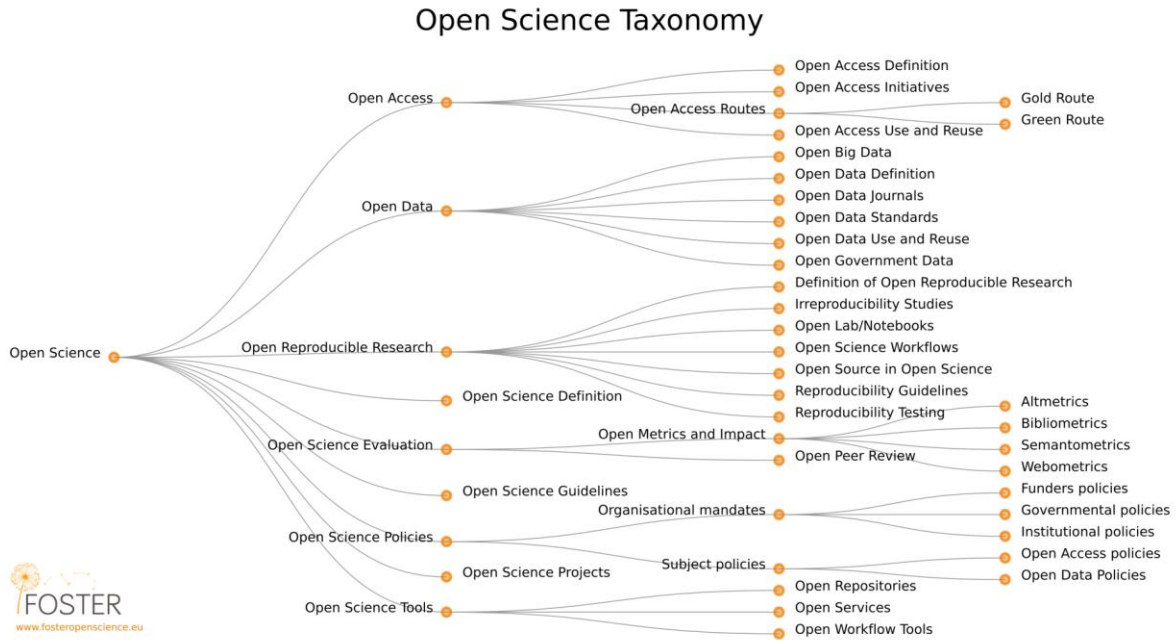
Open Notebooks - an emerging practice, documenting and sharing the experimental process of trial and error;
Open Data - managing research data in a way that optimises access, discoverability and sharing for use and re-use;
Open Research Software - documenting research code and routines, and making them freely accessible and available for collaboration;
Open Access - making all published outputs freely accessible for maximum use and impact.

In order to achieve this openness in science, each element of the research process should:

- **Be publicly available** - it is difficult to benefit and use knowledge hidden behind username and password barriers, or if it does not contain the right metadata to make it discoverable.

- **Be re-usable** - research outputs must be licensed appropriately so that prospective users know clearly any limitations on re-use.
- **Induce collaboration** between researchers through better access and better online tools;
- **Be transparent and have appropriate metadata** to provide clear statements of how research output was produced, and can be re-used.

A more concrete exposition of Open Science and its many branches is provided by the taxonomy developed by the European FP7 FOSTER project in support of its aim of putting in place "sustainable mechanisms for EU researchers to FOSTER OPEN SCIENCE in their daily workflow, thus supporting researchers optimizing their research visibility and impact, the adoption of EU open access policies." The taxonomy covers not only the constituent elements of OS but also supporting tools, measurements and mechanisms.



1.3 How open is our research?

Further evidence of this widening from Open Access and Open Data to Open Science is provided by the development by SPARC Europe of a tool for visualising, discussing and monitoring how open an institution's research is. It takes the form of a radar diagram generated by confirming status or actions, or estimating percentages, in eleven main topic areas:

The eleven topic areas are exhaustive and demonstrate the potential extent and complexity of Open Science, and the challenges faced by institutions and funders in bringing it about.

Such diagrams can be used in a number of ways, for instance as an assessment tool, to generate discussion and to inform policy-making.

2. Open Data

2.1 What is it?

As we noted at the start of this document, while Open Access to research outputs has a long history and development, Open Data has come into scope somewhat later. The OECD report already quoted makes the rationale specific (2):

> *"… reducing duplication and the costs of creating, transferring and re-using data; enabling more research on the same data; … increasing transparency and quality in the research validation process, by allowing greater replication and validation of scientific results."*

Research data can be defined simply as whatever is either produced in research or evidences research outputs.

The European Commission's definition is: "information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation". Examples include: statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings, images (3).

The 2012 European Commission's recommendation on access to and preservation of scientific information states that: "open access to scientific research data enhances data quality, reduces the need for duplication of research, speeds up scientific progress and helps to combat scientific fraud" (4). Unsurprisingly other funders also require open access to the data produced as a result of their funding. The Wellcome Trust for instance has been a leader in the field. Its *Policy on data management and sharing* (2010) states: "The Wellcome Trust expects all of its funded researchers to maximise the availability of research data with as few restrictions as possible."

Over time policies have developed. Commonly they will now include the following elements (5):

- **Timing**: when publication should take place;
- **Data plan**: requirements for a technical management plan;
- **Access and sharing**: what exactly will need to be available for public use;
- **Long term curation**: data creation and sustainability;
- **Monitoring**: any monitoring that will be carried out by the funding body and guidance available;
- **Storage**: details of the appropriate repository or data centre to be used;
- **Costs**: where costs can be claimed from and when.

Making data open is in some, particularly technical, senses more complex than making research outputs open: data collected must be capable of being verified, processed and re-used. However there are many resources covering all aspects of Open Data policies and practice now made available, for instance, by the Digital Curation Centre.

HEFCE, Research Councils UK (RCUK), Universities UK (UUK) and Wellcome recently (July 2016) published *Concordat On Open Research Data*, which is meant to:

" … [help] ensure that the research data gathered and generated by members of the UK research community is made openly available for use by others wherever possible in a manner consistent with relevant legal, ethical and regulatory frameworks and norms … The intention [of the Concordat] is to establish sound principles which respect the needs of all parties. It is not the intention to mandate, codify or require specific activities, but to establish a set of expectations of good practice with the intention of establishing open research data as the desired position for publicly-funded research over the long-term."

The development and promulgation of such principles is welcome. The *Concordat* also gives clear definitions and examples demonstrating that data are the result of humanities as well as scientific research:

"Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence… They may include, for example, statistics, collections of digital images, sound recordings,

transcripts of interviews, survey data and fieldwork observations with appropriate annotations, an interpretation, an artwork, archives, found objects, published texts or a manuscript."

The *Concordat* also recognises that:

"Not all research data can be open and the Concordat recognises that access may need to be managed in order to maintain confidentiality, guard against unreasonable cost, protect individuals' privacy, respect consent terms, as well as managing security or other risks."

In its *Guidelines* cited above, the European Commission takes as its Open Data mantra "as open as possible, as closed as necessary" and gives more detailed exemptions to data whose publication would be:

- Incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited;
- Incompatible with the need for confidentiality in connection with security issues;
- Incompatible with rules on protecting personal data.

FORCE11 has also published the FAIR data principles. Data should be:

- **Findable**: easy to find the data and the metadata for both humans and computers - persistent identifiers (PIDs);
- **Accessible**: data should be retrievable by their identifier using a standardised and open communications protocol;
- **Interoperable**: data should be able to be combined with and used with other data or tools. The format of the data should therefore be open and interpretable for various tools;
- **Re-usable**: metadata and data should be well described so that they can be replicated and/or combined in different settings.

2.2 What do researchers think of it?

Open Data may be in its infancy, but already there are outspoken champions among researchers. The views of fourteen researchers from seven different countries, active in diverse disciplines, were collected and published in 2017 by SPARC Europe. A number of themes emerge.

There are various prominent **rationales** for Open Data:

- Data produced as a result of public funding should be publicly available.
- It is only possible to validate or reproduce research findings if the underlying data and tools are available. Otherwise they have to be taken on trust.

- As is now very often the case, independent research groups around the world creating their own data gives rise to inefficiencies.
- Data can often be re-used, for instance being subject to different methodologies or coupled with other data.
- Open Data alone makes possible the creation of very large data sets, which can be exploited by machine techniques such as data mining.

The following actions are recommended to foster Open Data:

- There needs to be a change in research culture, so that sharing data becomes the norm. This of course depends on the motivation of researchers through academic incentives.
- Such cultural change may be speeded by gaining the active support of senior researchers and managers.
- Funders' policies can play a very significant role in achieving such change.
- Open Data (and Open Science) must become an integral part of researchers' education, not something separate.
- It must be made as easy as possible for researchers to deposit and share their data.

Some characteristics of an **Open Data world** are identified as:

- Knowledge creation will be accelerated, producing real-world benefits, particularly for medicine and business.
- It should be possible to draw on or incorporate large data sets created outside academia, for instance in transport, meteorology and medicine.
- The ready availability of data, with appropriate metadata, should drive the development of interdisciplinary research.

A very few downsides to Open Data were identified, including possible breaches of confidentiality and researchers' perception of the data they create as being their own property.

3. The Future

Open Science is moving centre-stage, with a rationale of improving efficiency in science and speeding the transfer of knowledge. We have seen a surge in practical developments for Open Data, mirroring the well-established repositories for research outputs, and in the development and application of model policies and principles.

It is very easy to become blinded by the interesting detail of these advances. However it is salutary to paraphrase Bill Clinton's mantra on the economy: "it's the research, stupid". It is researchers themselves, funders, governments, supra-national bodies such as the European Commission and industry and commerce that will benefit directly from and drive this openness. The benefits are potentially, huge, multiplying the return on investment in research, accelerating the research process and involving a full range of interested citizens.

We can already see new paradigms and structures arising. As information professionals we are already closely involved in their development. We must be seen to be leading the move to the new pervasive openness.

References

(1) OECD 2015.
(2) OECD 2015. Making Open Science a Reality.
(3) Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020
(4) European Commission 2012 recommendation on access to and preservation of scientific information.
(5) Guy, M. and Ploeger, L. 2015. PASTEUR4OA Briefing Paper: Open Access to Research Data.

Academic research and its relevance to enterprise search success

Martin White, Intranet Focus

martin.white@intranetfocus.com

At a time when maintaining a technology-employee balance is becoming increasingly difficult it is important to be able to take advantage of the substantial amount of academic research that is being undertaken on topics that include information retrieval, digital assistants, enterprise social networks, information systems adoption and collaboration effectiveness. I should say here that there is virtually no research on intranets. Academic research is certainly not being carried out only within universities; many excellent case studies have been published over the last few years. I have a collection of over 200research papers across the topics I've mentioned above, but I do have the benefit of access to the extensive digital resources of the University of Sheffield.

There are three challenges in using this research in practice. The first is finding the research papers. Google Scholar is my first choice because it offers date range search and also lists open source versions of published papers where they are available. Microsoft Academic and BASE (Bielefeld Academic Search Engine) also have their merits. The ACM Digital Library provides access to ACM journals and conference proceedings and IEEE Xplore does the same for the IEEE. Elsevier offers the Scopus database service for its own journals and books and there is also Web of Science from Clarivate. However, that leads on to the access problem. The ACM and IEEE databases are only available to members and Elsevier journals are behind a subscription pay wall. That is where Google Scholar is so useful in providing open source versions. If there appears not to be an open source equivalent putting the title of the research paper into Google as a search query sometimes highlights options that are not listed in Google Scholar.

Reading between the lines

The third challenge is in working out how to read an academic paper. My advice is to read the introduction and then jump to the bibliography at the end. The main purpose in doing this is to see what the date is of the most recently cited paper. If this is more than around three years old the chances are that the paper is the published version of a PhD thesis. Then move backwards to the section that gives an assessment of the extent to which the research scales, and the factors that might have influenced the outcomes of the research. This is usually a very honest assessment. One of the hot topics in academic research at present is "replicability". Too many papers publish results that other teams cannot emulate. Still working backwards you will come to the conclusions and a discussion of the research results. It can be helpful to read these final sections before reading all the

experimental methodology and outcomes. Without having a sense of the outcomes it is easy to become lost. The final step in the evaluation of the paper is to check on the authors to see how much previous work they have undertaken in the subject area of the paper.

In my experience people who see little value in academic research often do so on the basis of anecdotal evidence and not from personal experience. It is not easy to access this information (though it is usually possible to buy an individual paper) but the benefits can be quite considerable. Academic research may not provide definitive answers to very difficult issues but it can provide a vendor-independent framework for discussion and inspiration. In this article I am going to highlight some exceptionally valuable academic research that has been published over the last few months that has significant implications for enterprise search success.

Enterprise search satisfaction

Over the last two decades especially Microsoft Research has undertaken a significant amount of research into how people search on the web. One visible sign of this research effort is 'Interactions with Search Systems' by Ryen White, leader of the Cortana Research team at Microsoft. The bibliography lists over one thousand six hundred research papers but none of these addresses the way in which employees interact with enterprise search applications. The reason for this is that there have been no papers published on the topic except for a few examples that take a very narrow view of the topic. A number of search consultants have good anecdotal stories to tell about enterprise search behaviours but they are not based on anything approaching a rigorous base of sampling and analysis. As a result there are potential dangers in extrapolating results from web search and applying them to enterprise research.

Now at last Dr Paul Cleverley and Professor Simon Burnett (Robert Gordon University) have published (in the Journal of Information Science) what is without doubt a landmark research paper on the factors that influence user satisfaction with enterprise search applications. The Journal of Information Science is a subscription research journal published by Sage but there is an open access Author Accepted version on OpenAir@RGU, which is the open access repository of Robert Gordon University, Aberdeen.

There are three reasons for me applying the "landmark" label to this paper. The first is the scale, with more than one thousand users in a large multinational company providing feedback over a period of two years. Nothing on this scale has ever been undertaken. Over the last sixteen years the company has implemented three different search applications.

The second reason is that Paul Cleverley is a geophysicist who moved into information science roles in the oil and gas industry and then last year was awarded a PhD for his work on the use of filters and facets in enterprise search. So here is a discipline expert with a very solid understanding of research methodologies applying all his experience and expertise to understanding enterprise search behaviours. Moreover, since six out of the

ten largest companies in the world are in this sector, there is at least a reasonable expectation that the outcomes will be similar in other large multi-national companies.

The third reason is that this paper proves that it is possible for research with an academic rigour to be undertaken within an organisation. Academics come up with all sorts of reasons for not attempting research within organisations – now this paper and its methodology shows that it can be done, and how it can be done. Hopefully others will now follow this lead.

The methodology is what is usually referred to as a longitudinal mixed methods approach. First feedback was obtained from the search user-interface to gauge satisfaction with the search outcomes. Second interviews were carried out with members of the thirteen internal and contract staff supporting the search application. The two data sets were then triangulated to highlight areas of agreement (all but two), dissonance (none) and silence (two). The study was longitudinal, with the same group of users being monitored over a period of two years. The interviews were coded so that a clear differentiation could be created between satisfaction and dissatisfaction. The areas of dissatisfaction were Technology Factors, Information Factors and Literacy Factors. The analysis of the outcomes of the research is very thorough, and the paper closes with a definitive bibliography of almost one hundred and fifty research papers, reports and books.

Trying to summarise the outcomes of a twenty-four-page research paper is very difficult. Nevertheless, there are some outcomes from the research project that stand out as being especially worthy of close attention by the enterprise search community. In commenting on them in this article I run the danger of taking them out of context of the research and analysis so I would encourage you to read the full paper.

In the paper the factors identified that predominately influenced satisfaction were technology, information quality, information literacy and task utility. The technology factors include search tool reliability, search ranking and query syntax handling. In total these factors were the largest single group (38%) and that could be used as a justification for investing further in search technology. However together information factors (36%) and literacy factors (26%) accounted for 62% of the reasons for dissatisfaction and to me that indicates that technology investment on its own is not going to make a significant difference to search satisfaction.

Moving on to search-level metrics, the search application was used by around seventy thousand staff each month and generated over four hundred and fifty thousand search queries. The average query length was 1.89 words and the top thirty most frequent searches fell from 14% of all search queries at the start of the project to just 8% at the end of the project two years later when of course users had gained substantially more experience with the application. This confirms anecdotal evidence that the tail of low frequency queries is very long in the enterprise environment. In my view this has significant implications for “cognitive search” because there will be such low levels of use data from the majority of the queries to be able to predict optimal results. The percentage of results with “no results” decreased from 0.4% to 0.3% over the same period.

Throughout the section in which verbatim comments are included it is very clear that resolving these comments requires knowledge of the technology, the content and the use case. This requires a skilled search support team, but from the outcomes of the research this team also has an important role to play in overcoming the problems of a lack of search literacy. The literacy factors include the selection of a query term, results scanning and familiarity with the search application. As over 25% of all dissatisfaction events were attributed to poor search literacy there has to be a major question mark over the view that when technology and information content are optimised search outcomes will take care of themselves.

My final comment is so well stated by the authors that I will reproduce it from the paper.

"The importance of configuration in enterprise search was evident, where an unintentional change biasing documents over web pages let to sub-optimal results. With an average query length of approximately two words, made by users to dynamic growing corpus sizes, it is unlikely that many information needs will be met without constant configuration, promotion of authoritative (trusted) corporate information and monitoring of performance."

With all research projects there is always the danger that the outcomes are not scalable and extensible to other organisations. Reading this paper, so many of the comments aligned with the experiences I have had with probably one hundred or more enterprise search-related projects that I have every confidence that the outcomes will translate (with due care and attention) to any organisation that depends on being able to offer complete and effective access to global information repositories.

No matter how large your organisation, if you have responsibility for search management you should be taking this remarkable paper, marking it up paragraph by paragraph, and then using it to benchmark your approach to achieving the levels of search satisfaction that your employees expect.

The importance of Search Results Page (SERP) formatting

Over the years I have been involved in many usability tests where employees are given a search task to perform, such as, "Find the technical support manager for air compressors in Argentina." While an apparently simple task, the diversity of approaches employees take becomes visible very quickly. Multiple start points are immediately apparent, reflecting the experience and expertise of the employee performing the search.

An enormous amount of research has gone into information seeking over the last few decades. A survey of this research published in 2007 ran to over four hundred pages, and the pace has accelerated since then.

When assessing enterprise search performance, the focus is always on counting clicks, worrying about "precision at k," mean reciprocal rank, and other formulae that assume

users work their way sequentially through the ranked list of results. These clicks do not reveal an element of the search process: the stopping strategy for the search.

Relatively little research has been carried out into what might cause a user to stop a search. In enterprise search this could be something as simple as the date shown in the result snippet. One user may decide anything older than 2016 is not going to be relevant, while another user may stop at 2017.

Click traffic will not make this stopping strategy apparent, especially in cases where a session is halted and then resumed with a different query some time later.

The scent of a SERP

Peter Pirolli and Stuart Card developed the information foraging model for information seeking while working at Xerox PARC in the early 1990s. Ed Chi, a fellow Xerox PARC employee, further developed the model in the late '90s. The concept of an “information scent” refers to the way (for example) pigs can find truffles even though they are well hidden.

So what's the connection between truffle hunting and Search Engine Results Pages (SERPs)? The answer is a search user's view of results pages is informed by a wide range of proximal clues, which together create an information scent in the mind of the searcher. For example, a glance at ten PowerPoint files listed on the first page of results could bring a search to an abrupt halt before it has even started.

David Maxwell, a PhD student in computer science at the University of Glasgow and Leif Azzopardi, associate professor at University of Strathclyde, presented a paper at the 40th European Conference on Information Retrieval in March, which prompted this column. You can download the paper (along with many other interesting papers) from Maxwell’s personal website. In their paper, Maxwell and Azzopardi hypothesize, model and then validate the impact the information scent of a SERP has on stopping strategies and therefore, search performance. In summary (and there is a substantial amount of data and analysis in the paper), they believe the role the quality of SERP presentation has had on search effectiveness and satisfaction has been significantly underestimated.

The paper goes on to discuss the search ability of users. Again, in the "click count" world, all users are assumed to have equal search proficiency and an equal command of the languages being used on the SERP. The paper shows search proficiency influences opinions about the usefulness of the page based on information clues from SERP, and the authors set out some potential categories of user proficiency. Another paper by Leif Azzopardi, this time with Paul Thomas and Nick Craswell (both eminent members of Microsoft Research) takes up the SERP topic. It is entitled Measuring the Utility of Search Engine Result Pages: An Information Foraging Based Measure and can be downloaded from the Microsoft Research site. A SERP is typically constructed from:

- A header, where the query box and query statistics are displayed.
- The core, where the main set of algorithmic results is shown along with advertisements and other answers e.g. navigational entity cards, image, video and news elements.
- Often, a right rail, where entity cards, advertisements, related searches, etc. are shown.
- A footer with navigational cues to the next or previous page.

Now this of course is a typical web site SERP but the principles apply to enterprise search applications as well. The essence of their argument is that traditional approaches to search metrics treat each result in isolation. Their research suggests that as a user works their way down a set of results they learn from the process and there is a cumulative effect that may lead them to take a different perspective on the rank order. This affects stopping strategies and the quality of information snippets in results lists may also be having a significant impact on the user's assessment of the results. The end result is that all the work that has been carried out in "precision at n" may not be a good indicator of search performance.

Implications for Enterprise Search

As with any research, the outcomes presented in this paper should not be generalized without carefully considering the methodology and analysis. The authors rightly set out where further research is required to understand more clearly the impact of information scent on stopping point determination. This research will undoubtedly lead to a more reliable assessment of information seeking behaviours in an organisation.

Even so, I believe all enterprise search managers can take away some lessons from the current research:

- Relying only on search click traffic analysis is rather like assessing a holiday beach from a monochrome print.
- Usability studies provide essential information about how the user is performing, not just how the system is performing.
- SERP presentation values are likely to have a significant impact on achieving high levels of search satisfaction. Further research (at an organisation level) will be needed to assess the improvement in performance.
- If this proves to be the case, then using cognitive search applications to present a small number of highly personalized results could be counter-productive.
- Key performance indicators, such as "precision at k" calculations, may potentially need to be completely reconsidered.

There are many other examples where academics are working on real-world problems. Earlier this year I wrote three posts about research into enterprise social networks, a topic on which there is a lot of passion but (in my view) very little insight. For the last few years I have been writing a Perspectives column for Business Information Review in which I scan through all the Sage journals to find academic research that could be of interest to

information and knowledge managers and then write a summary that conveys the potential impact of the research on organisations. Each column takes about a week to research and write but it (like eLucidate) is a task I enjoy, and from which I have gained many insights into novel approaches to the organisational management of information and knowledge.

Can a Library Management System (LMS) be used to archive and preserve your social media tweets?

Marlize Palmer, Head of Information and Archive Services, Welsh Government

James Dawes-Hughes, Information and Archive Services, Welsh Government

marlize.palmer@gov.wales

Working with our LMS provider, Soutron[1], the Welsh Government's Information (Library) and Archive Service undertook a proof of concept and/or mini-pilot to make the Welsh Government's "Organ Donation" tweets available via our Library Management System (LMS).

Twitter is a 'micro blogging' platform that allows users to post short text messages up to 280 characters in length and converse with other users via their phones or web browsers. Unlike email or text messaging on mobile phones, these conversations take place publicly, unless a user's tweets are protected. Twitter is experiencing a phenomenal adoption curve in the UK and is being used increasingly by government departments, Members of Parliament, a number of our stakeholders, as well as millions of businesses, non-government organisations and individuals. It is free to use with a relatively low impact on resources and has the potential to deliver many benefits in support of the Welsh Government's communications objectives.

The Welsh Government uses a significant number of separate Twitters feeds on a day-to-day basis, but the three main feeds used are:

- @Welsh Government — The English language account
- @LlywodraethCym — The Welsh language account
- @FMWales — The First Minister of Wales' account

[1] https://www.soutron.com/
https://www.soutron.com/flexible-database-expanded-reach/

These three accounts contain over twenty thousand separate tweets and provide an overview of the work undertaken by the Welsh Government since the organisation started tweeting in May 2009.

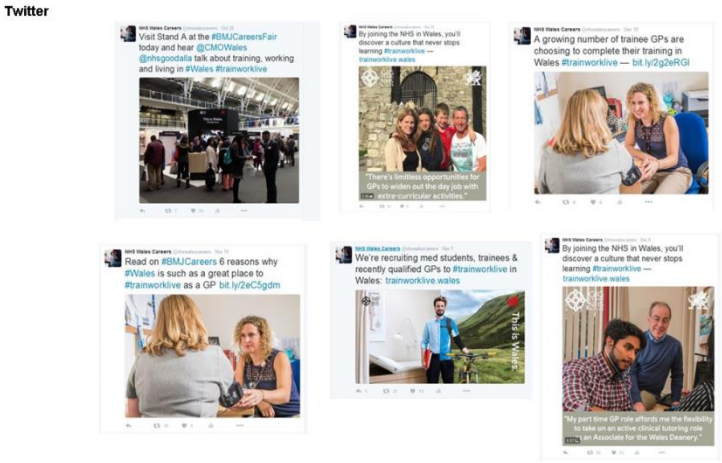
Capturing and archiving social media for re-use is a major challenge. Making the data (tweets) searchable and interpretable is challenging, but it's a necessity, especially when considering that most people use Twitter as one of their main sources of news, information and communication. It is important to find ways of incorporating social media into knowledge structures and archives, as they are valuable resources. Social media records can be used for research, especially when used in context with other digital resources such as email and Word documents (i.e. "linked information").

Tweets could also enrich our collection of consultation documents and provide additional context to the collections.

Background

On the 1st of December 2015, the Human Transplantation (Wales) Act came into full effect. It introduced a soft opt-out system for consent to organ and tissue donation[1]. The Welsh Government (WG) used Twitter extensively to promote its PR campaign to support Organ Donation Wales during a six-month period (1st of June – 30th of November 2016) to increase organ donation registrations, attitude and awareness.

The campaign comprised of weekly schedules of social content (two Facebook posts and one tweet per day) to support monthly themes including: Living Donors; 18-34 year olds; Time to Talk / Organ Donation Week (over 520 posts & tweets in total). It is estimated that the Welsh Government managed to reach a total audience of around 1.5 million via its "Organ Donation Wales" social media campaign.



The collection

We decided to trial hosting approximately one hundred and seventy eight tweets and their associated metadata on our LMS. The tweets that we identified for capturing on our LMS were held in the following file formats:

1) A PDF screenshot of the tweet in question.
2) A txt file containing just the text from the tweet in question.
3) An HTML file including the metadata of the tweet in question.

In addition to this, the team used The National Archives' (TNA) file profiling tool DROID[2] to help profile the file formats in the collection, and to create a simple CSV metadata spread sheet consisting of a file name, short description and type of tweet – see example image.

File Name	Description	Type	Transfer
ESIV-F-4-00000001.pdf	6 mis tan Rhoi Organau	Retweet	Yes
ESIV-F-4-00000002.pdf	100 diwrnod tan Rhoi Organau	Tweet	Yes
ESIV-F-4-00000003.pdf	Datganiad gan Gweinidog Iechyd	Retweet	Yes
ESIV-F-4-00000004.pdf	Rhoi Organau wedi achub bywydau	Tweet	Yes
ESIV-F-4-00000005.pdf	74% yn gwybod am y system Rhoi Organau	Tweet	Yes
ESIV-F-4-00000006.pdf	Oes gennych chi gwestiwn am Rhoi Organau?	Retweet	Yes
ESIV-F-4-00000007.pdf	Statement from Health Minister	Retweet	Yes
ESIV-F-4-00000008.pdf	Paratoadau terfynal ar gyfer Rhoi Organau	Tweet	Yes
ESIV-F-4-00000009.pdf	6 mis tan Rhoi Organau	Retweet	Yes
ESIV-F-4-00000010.pdf	Y system Rhoi Organau Newydd	Tweet	Yes
ESIV-F-4-00000012.pdf	Full scrape of @LlywodraethCym	Feed	No
ESIV-F-4-00000013.pdf	Three tweets with #RhoddwrOrganau	Hashtag	No
ESIV-F-4-00000014.pdf	Bywydau wedi'u hachub	Tweet	Yes
ESIV-F-4-00000015.pdf	Paratoadau terfynal ar gyfer Rhoi Organau	Tweet	Yes
ESIV-F-4-00000016.pdf	Full scrape of @fmwales	Feed	No

Complexities and Challenges

The data itself, privacy concerns and the need for us to develop appropriate access and usage policies were some of the complexities and challenges we faced. We explored these issues and the following as part of our mini-pilot:

- How to process and organise these tweets as well as how to physically store them.
- How to provide useful means of access and retrieval.
- Challenges involving policy, such as the creation of appropriate access controls to the collection, whether any information should be censored or restricted (we currently envisage that only WG staff will have access to these tweets).
- What is the minimum/maximum amount of metadata that accompanies each tweet that should be captured? [3]
- Is our LMS, Soutron, capable of providing access to these tweets – what are the constraints regarding providing direct access to the data elements within the Twitter Archive, or with any restrictions?
- How much work is involved "processing" tweets as part of a broader collection and what do these changes mean to our current understanding of a record or library item? (i.e. What type of indexing and processing is required by information professionals and/or technology experts to make the collection accessible and re-usable.)
- Moving to second generation digital archiving, what type of sophisticated access tools might be required to provide a "basic level of access" for researchers and users of the collection?
- Broader ethical considerations of the very existence of such a collection (i.e. should there be any access and/or content restrictions and if so would a time limited "take-down" policy be sufficient?)
- Are there any privacy concerns about creating a permanent archive of government tweets and are there any GDPR related issues?
- Should we do this at all (i.e. will we be preserving a mountain of worthless information?)

Copyright issues

Copyright protection exists at the moment of creation and it is fixed in a tangible form. But copyright does not protect facts, ideas, systems, or methods of operation, although it may protect the way these things are expressed.

The general consensus seems to be that most tweets are falling outside copyright because very few are copyrightable or may rise to the level of copyright protection. Nevertheless, we decided not to capture other people's reply tweets to Welsh Government's original tweets for the pilot.

The Data

We worked with Hanzo to extract our previously harvested and archived Twitter presence between 2014-2016. This involved technicians at Hanzo writing a code to extract all the relevant tweets.

Four separate searches were undertaken using the following search terms.

- organ donation
- rhoi organau (Welsh language term)
- #organdonation
- #rhoiorganau

Once identified, Hanzo provided the data via download URLs that would allow us to transfer the collection to our systems using Winzip. The data consisted of five different types.

Data Types	
Tweets	Tweets produced by a Welsh Government account regarding organ donations
Retweets	Tweets retweeted by a Welsh Government account regarding organ donations
Feeds	These were screenshots of entire Twitter accounts
Hashtags	These were screenshots of the search results from Twitter using a particular hashtag relevant to organ donation
Replies	Replies from other accounts to Welsh Government Tweets regarding organ donation

After analysing the data (the five separate types shown in the image), we decided that the majority of Tweets and Retweets were worthy of retention, whilst the Feeds, Hash tags and Replies were less likely to have any research or re-use value.

Storage of tweets for research and re-use on LMS

The next step was to work with our Library Management System (LMS) provider, Soutron to upload the collection of tweets onto our LMS to enable us to preserve and re-use them for research purposes. We required our LMS to store and provide access to tweets irrespective of whether they were saved as attachments, images or links and to store the metadata accompanying each tweet.

Soutron started by examining the different variants of metadata and PDFs that were being harvested from the selected Welsh Government's Twitter accounts. The first thing that stood out was that the PDF contained only an image. It didn't include any metadata which, if it had, they then would have been able to automatically index the metadata on

upload of the image. Soutron's index is immediately updated and the search would find the image's content. However, a "text" file was also provided with the PDF file, using a common name, which meant that they could load this alongside the PDF as part of an automated load function. The idea of manually typing the words of each tweet into the LMS would have made the project unworkable.

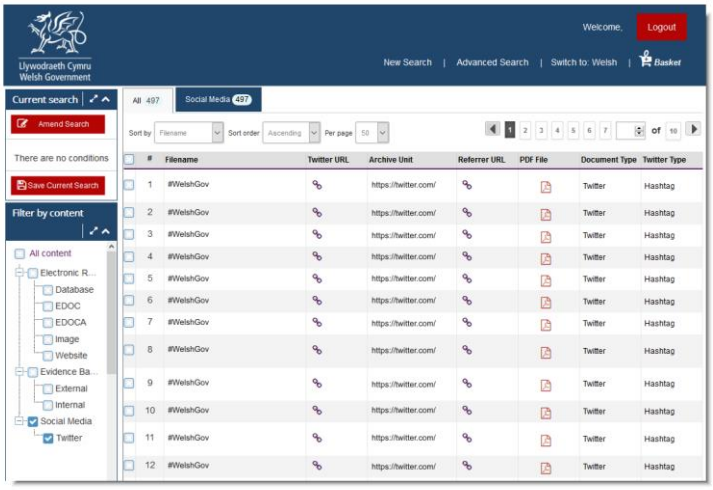
There was additional metadata that accompanied the Twitter text file and image file, but it was in html format, not ideal to work with. They were able to quickly clean this metadata and load it into Excel so that this could be done. This extra bit of data was important as it contained the URLs of the tweet as well as the referral URL. Critically and very usefully, it contained the name of the PDF file containing the image of the tweet, which meant that they were able to link the metadata with the PDF using the standard Soutron importer tool.

In addition, the simple CSV created by the Welsh Government's Information and Archive Services with the language and description of the tweet provided Soutron with further metadata that could be used for the collection.

We jointly determined the fields that were required and created the required record and field structure using the standard menu driven facilities in Soutron that control the database structure. This included customising a new dedicated Search Result template for tweets.

Testing the Results

A test database was used to set up these tasks and to load initial test data sets. The URL to the test system was provided to us to evaluate. The result was very positive and the original objectives were all achieved.



Welsh Government LMS Screenshot of the Soutron LMS Application

It became clear by experimenting and working with this data set that there is yet further potential. Using Twitter's API's it may be possible to automatically extract and index data from specific Twitter accounts, to archive important data that otherwise may get lost.
It is exciting to capture and preserve data in this new medium. Librarians and Information Professionals have always been at the forefront of technologies and play a pivotal role to manage vital information, more and more of which is outside of traditional print material.
To explore how Soutron can use their LMS to archive your social media content, get in touch with them today at www.soutron.com.

Notes:

[1] The Human Transplantation (Wales) Act 2013 (anaw 5) (Welsh: Deddf Trawsblannu Dynol (Cymru) 2013) is an act of the National Assembly for Wales, passed in July 2013. It permits an opt-out system of organ donation, known as presumed consent, or deemed consent. The Act allows hospitals to presume that people aged 18 or over, who have been resident in Wales for over 12 months, want to donate their organs at their death, unless they have specifically objected. The Act varies the Law of England and Wales in Wales (still applicable in England), which relied on an opt-in system; whereby only those who have signed the NHS organ donation register, or whose families agreed, were considered to have consented to be organ donors.

[2] http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/

[3] More than just the plain text that a user types into the Twitter interface, each tweet contains 150 pieces of metadata, such as a unique numerical ID, a timestamp, a location stamp, IDs for any replies, favourites and retweets that the tweet gets, the language, the date the account was created, the URL of the author if a Web site is referenced, the number of followers, and numerous other technical specifications.

Online Resources

Joy Cadwallader, Aberystwyth University (Aberystwyth Online User Group)

Please send your submissions for the next issue to jrc@aber.ac.uk

BBC

The BBC has made over sixteen thousand sound effects freely available online for personal, educational or research use. *BBC Sound Effects* (currently in beta) is a great resource with everything from the sound of a V2 rocket launch to a cat's miaow available for download in .wav format. The service is delivered via the Research and Education Space (RES) which (from their FAQ), "began as a partnership project between the BBC, Jisc and [BUFVC] Learning on Screen … [which] has now come to an end, however the BBC will continue to work with public partners to develop RES as an open platform for publishing linked open data."

Clarivate Analytics/Kopernio

After receiving development support and funding from the UK government support, as recently as February 2018 (via Innovate UK), the AI-technology start-up Kopernio has now been bought out by Clarivate Analytics (CA). CA say this is, "to create the definitive publisher-neutral platform for research workflow and analysis for scientific researchers, publishers and institutions worldwide". Using AI, Kopernio intend that their product delivers a legal one-click service to articles and research papers regardless of their location e.g. subscribed to journal, pre-print server, repository, blog etc., or your location i.e. on or off-site/off-campus, integrates with GoogleScholar and PubMed, and stores what you have found for later. The app is available now as a free download. CA intend to integrate it with their other services and say, "Kopernio intends to quickly build a valuable commercial offering for publishers and academic institutions".

Dublin City University (DCU)

In a partnership with DCU's Office of the Vice President of Research & Innovation and DCU Libraries, Dublin City University has announced the launch of DCU Press which is the first open access university press in Ireland. The press release also explains how the DCU Press, "is a leading innovation in line with the ambitions of Horizon Europe", the European Commission's newly-adopted €100 billion research and innovation framework programme.

Elsevier

Florida State University Libraries (FSU) are to drop their Elsevier "big deal" from January 2019. In an announcement by Julia Zimmerman (Dean of University Libraries) FSU are

paying two million dollars per annum with a 4% annual rise, which is more than neighbouring universities due to, "a poorly thought-out twenty-year-old contract between Elsevier and the State University System". FSU have been unable to extricate themselves from the contract so will instead subscribe to a subset of titles, use interlibrary loans and, "expedited delivery within minutes ($30, subsidized by the library)". Julia also states that they will be able to buy other materials previously requested but refused, "because so much of our budget has been consumed by Elsevier". Florida State is not the first to take the plunge; this useful article in *Inside Higher Ed* by Lindsay McKenzie provides a timely recap of "big deal" cancellations and diverse views on these actions. In July many researchers in Sweden and Germany lost access to newly-published articles in Elsevier journals after negotiations over subscription contracts broke down.

Gale

Gale have launched a new online archive of primary source material: *Political Extremism & Radicalism in the Twentieth Century: Far-right and Left Political Groups in the U.S., Europe and Australia* comprises, "a diverse range of content, including campaign materials, propaganda, government records and various ephemera". Content from the UK is drawn from the National Archives and Searchlight archive (based at the University of Northampton), "an information service that aims to expose racist and fascist groups". Subscribers will have access to, "more than six hundred thousand pages of content and more than 42 audio histories with full transcripts."

Google

Google has developed a new freely available semantic search tool called *Talk to Books*, which uses AI to return answers from over one hundred thousand books to natural language enquiries. The "learn more" page explains that, "The input data is a billion pairs of statements, where the second statement is a response to the first one", and, "Once the AI has learned from that data, it is then able to predict how likely one statement would follow another as a response". I gave it a try and it felt powerful if a bit gimmicky but at first the results appeared to be all from non-fiction source books. So I tried searching for '*should Elizabeth Bennet marry Mr Darcy*' and found an answer from *Pride and Prejudice* in the top three results, with extracts from academic works including *Kantian Ethics* and *Jane Austen, Game Theorist* in the top ten results.

IFLA

IFLA have launched a new feature called Country Pages to their Library Map of the World. Once completed, users of the map will be able to select any country and gain an overview of their library provision including governance, funding modules, policies and their national library. It's early days with just a few pages ready so far. However, I clicked on the Countries tab and found a detailed page on Libraries in Latvia. I discovered that all Latvian libraries undergo accreditation every five years (under their Library Act) and that the National Library of Latvia won, "Library of the Year Award in the International Excellence Awards at the London Book Fair in 2018".

Irish Film Institute

The Irish Film Institute has released *The Irish Independence Film Collection*, a set of freely available newsreel material filmed between 1900 and 1930; a period including such major events as the Easter Rising, the Irish War of Independence and the Irish Civil War. Footage includes Michael Collins speaking to huge crowds in Dublin in 1922 following the independence treaty of 1921, a reel on the trial of Roger Casement for treason in 1916, and the re-opening of the Dublin GPO in 1929, which had been the headquarters for the leaders of the Easter Rising and was all but destroyed. The press release explains how the collection was digitised from a selection of British Pathé and British Film Institute archives' original nitrate films, which had been repatriated to Ireland for the first time as part of a Centenary of Commemoration project with funding from the Department of Culture, Heritage and the Gaeltacht.

Library of Congress

April 2018 was a busy month for eye-catching announcements from the Library of Congress. First a personal favourite, the release of additional content online from the Leonard Bernstein archives including *West Side Story* scene and musical sequence outlines, draft scripts, song lists, lyrics, audition notes etc., and later materials related to his involvement with the civil rights movement.

Next a proposal for a new mandate to deposit a copy of every ebook published in the US, regardless of whether or not it is also available in print, if requested by the Library. If approved this would also cover self-published ebooks. Thanks to Infodocket for the story at FCW.

Finally, Benjamin Franklin's papers have been digitised and made available online. As might be expected there are papers and letters of great historical importance here, relating to his roles as diplomat and statesmen, but also his scientific observations and correspondence. These include, "Franklin's scientific speculation on the speed of ships in 1775 while on board a vessel returning from England to America just before the Revolutionary War", his, "drawing of bifocal glasses, which he is credited with inventing", and his, "letter explaining the effects of lightning on a church steeple".

Nature

A new Nature Publishing Group journal about machine learning due for publication in January 2019 has been hit by a boycott of more than 2,500 machine learning researchers. It appears that research published in this strand of science is traditionally open access but *Nature Machine Intelligence* content will only be available by subscription. The signatories of the boycott will refuse to submit to, review or edit for the journal. A post on the *Retraction Watch* blog points to a significant footnote in recent history included in the petition: " … in 2001, the editorial board resigned from a subscription-based journal, *Machine Learning*, now published by Springer (which has merged with the publisher that

includes Nature Publishing Group), and formed the open-access journal *the Journal of Machine Learning Research* (JMLR)."

University of Glasgow

A set of forty two audio recordings in Gaelic and English of crofters, farm workers and fishermen made in the 1970s have been added to the online Audio Archive at the Digital Archive of Scottish Gaelic (DASG) by the University of Glasgow. The recordings have been digitised from reel-to-reel tapes and audiocassettes, and they, "will be fully transcribed and searchable with detailed contents". The Audio Archive *Cluas ri Claisneachd* comprises recordings made during the collection process for the Historical Dictionary of Scottish Gaelic and additional donations, and the Mòthan Archive of recordings made on Scottish islands in the 1990s. This announcement has come as DASG enjoys its 10th anniversary.

Looking to the future: Internet Librarian International 2018 A 20th birthday preview

Katherine Allen, Business Development Director, Information Today Europe

kat.allen@infotoday.com

Internet Librarian International (ILI) will take place at Olympia, London on the 16th - 17th October, with workshops on the 15th October. UKeiG members benefit from a 25% discount on the full conference fee.

ILI is celebrating its twentieth birthday this year. When the conference began in 1999, the first dot com bubble was in full swing and sci-fi film *The Matrix* was released – but few could have predicted the incredible pace of change in information sharing and seeking behaviours that was to come. Back then, it was a novelty to be able to buy mobile phones in supermarkets, and SMS texting was just taking off. Google and Amazon were established, but not yet the global behemoths they are today. YouTube, Facebook, Wikipedia, smart phones and tablets were several years away and open access was yet to become a reality.

In putting together this year's conference programme, we've been mindful of this incredible pace of change – but the conference theme "Celebrating innovative libraries and info professionals" looks to the future, by sharing the many ways in which libraries and information professionals from all sectors are creating ground-breaking and imaginative services that deliver real impact.

Cultivating knowledge communities

ILI's keynote speakers will each contribute to this year's theme. Opening keynoter Katherine Skinner is the Executive Director of the Educopia Institute, a US-based non-profit run by a small group of individuals with a passion for building communities, connecting like-minded people, and using collective action to advance libraries, archives, museums, and publishers. In her keynote "Cultivating knowledge communities", Katherine will share stories about how impact becomes magnified when institutions band together – and encourage delegates to align our actions to make system-level changes that favour knowledge and memory.

Fantastic future? Predicting promise and peril

ILI's day two keynote will come from Martin Hamilton, the resident futurist at Jisc, the organisation that provides digital solutions for UK education and research. Martin's job is to keep an eye open for emerging trends and new technologies, and see what can be done to exploit and embrace them – or to mitigate against them. In his keynote, Martin will be taking an in-depth look at some of the major trends in digital technology - including AI, blockchain, data driven decision-making, and open research practice. How will they impact library customers, organisations and information work? And what's emerging from the research lab that will change knowledge and information services in the future?

Six tracks highlight new roles, new skills, new tools, and changing services

This year for the first time, ILI offers a track called "Inclusion and inspiration: libraries making a difference" and the important theme of inclusivity, diversity and equality features throughout the conference. Tech-led innovations feature throughout, and delegates can move freely between any of the six tracks:

FUTURE FOCUS: THE NEXT-GEN LIBRARY, THE NEXT-GEN LIBRARIAN

New roles, new skillsets and new tools that librarians are using to ensure they stay at the cutting edge – and how the profession itself can attract, support and nurture a diverse membership.

UNDERSTANDING USERS, USAGE AND UX
Looking at user driven change and how libraries are working with users to understand their behaviour and design services that meet – or exceed – users' requirements.

INCLUSION AND INSPIRATION: LIBRARIES MAKING A DIFFERENCE
How are libraries and information professionals supporting social inclusion, engagement and equality?

CONTENT, COLLECTIONS, COLLABORATIONS
Collections, collaborations and curation, working with new formats and new audiences.

MAGICAL MARKETING
How are libraries influencing use and expanding audiences for their services and collections?

NEW SCHOLARLY COMMUNICATIONS
How and why the scholarly communications landscape is changing, and what this means for libraries and information professionals.

In addition, there are informal activities for all at ILI Extra, including one hour mini-workshops, a pop-up "Glass Room" workshop, live podcasting, and an invitation-only workshop specially designed to give new information professionals the chance to join in with the ILI experience.

So, while much has changed since the first conference twenty years ago, that pace of change can be expected to continue, driven by disruptive technology, new business models for information services, rapidly evolving user expectations, and political change. Over the years, we've seen librarians and information professionals rise to that quadruple challenge in many ways, large and small. It's been incredibly exciting to see so many inspiring and transformative projects showcased at the conference and we look forward to ILI delegates sharing many more innovations this year, and into the future.

Discounts for UKeIG Members

UKeIG members are entitled to claim a 25% discount on fees for the main conference using code UKEIG25 when registering. Discounts are also available for multiple delegates from the same organisation.

Internet Librarian International 2018
The Library Innovation Conference
16th & 17th October 2018, Workshops on 15th October 2018
Olympia, London
www.internet-librarian.com

ILI is co-located with Taxonomy Boot Camp London, now in its third year:
www.taxonomybootcamp.com/London
Further information from organisers, Information Today:
E: info@internet-librarian.com

The Strix Award and 2018 Annual Memorial Lecture: In Honour of Tony Kent's Life & Achievements

UKeiG has announced that the 4th Tony Kent Strix Annual Memorial Lecture 2018 is to be delivered by Maarten de Rijke, Professor of Computer Science at the University of Amsterdam, and will take place on the afternoon of Friday, 23rd November 2018 at The Geological Society, Burlington House, Piccadilly, London. Book here.

Last year UKeiG, in partnership with the International Society for Knowledge Organisation UK (ISKO UK), the Royal Society of Chemistry Chemical Information and Computer Applications Group (RSC CICAG) and the British Computer Society Information Retrieval Specialist Group (BCS IRSG) was delighted to announce that the winner of the prestigious Tony Kent Strix Award for 2017 was Professor de Rijke. The Award was presented to him by Doug Veal (Strix Chair) and David Ball (UKeiG Chair) in London on Friday October 20th 2017 in recognition of his major and sustained contributions to the field of information retrieval and web searching.

Professor de Rijke is a well-known and highly respected member of the international information retrieval community having made considerable and widely recognised contributions to the field. He has an impressive and formidable high impact publications record in a range of areas including semantic search, semi-structured retrieval and social media. He has produced influential research outputs on the large-scale semantic analysis of online content and on the analysis of subjective aspects of information (sentiment, credibility, memory, reputation and experiences). His contributions to information retrieval, in particular to the fast evolving areas of computational methods for analysing, understanding and enabling effective human interaction with information sources, have been profound.

He leads the Information and Language Processing Systems Group at the Informatics Institute of the University of Amsterdam. It is one of the world's leading academic research groups in information retrieval and intelligent information access, with projects on self-learning search engines, semantic search and the interface between information retrieval and artificial intelligence.

His 2018 Strix Lecture is entitled: 'Retrieval as Interaction.'

Abstract: Information retrieval systems, such as search engines, recommender systems and conversational agents, may well be the prime example of interactive systems to which people are exposed. Their development is best thought of as a two-stage process: off-line development followed by continued online adaptation based on interactions with users.

Off-line development, which involves evaluation and learning from annotated datasets or from logged interactions, is risk free in the sense that the system does not require the ability to make interactive interventions. In contrast, in online development retrieval systems use interactions and interventions for evaluation and for learning.

In the lecture Professor de Rijke will compare the off-line and online development phases. How much can a retrieval system learn off-line from historical interaction data? How much can it gain by being able to make interventions and explore new actions? Can we give bounds on the risks a retrieval system takes when performing online interventions? He will also discuss another set of questions that come up in this space and that have to do with how we should design systems that learn and adapt online and are respectful of their users.

*** This is a FREE event, open to everyone, BUT advance bookings ARE required ***

Full programme details:

- 1.30 Registration
- 2.00 Douglas Veal - Chairman's welcome
- 2.10 - Stella Dextre Clarke, 2006 Winner of the Tony Kent Strix Award, reflects on 'Then and Now: Contrasts in the retrieval environment.' Now retired from consultancy in information management, she is still active as Vice-Chair of the UK Chapter of ISKO (International society for knowledge Organization).
- 2.45 - Questions & Discussion
- 3.00 - Tea & coffee
- 3.45 - The Tony Kent Strix Annual Memorial Lecture
- 4.30 Questions & discussion
- 5.00 Meeting closes

Call for 2018 Strix Award nominations

The Tony Kent Strix Award was inaugurated in 1998 by the Institute of Information Scientists. UKeiG is now seeking nominations for 2018. The prestigious award is given in recognition of an outstanding practical innovation or achievement in the field of information retrieval in its widest sense, including search and data mining, for example. This could take the form of an application or service, or an overall appreciation of past achievements from which significant advances have emanated. The award is open to individuals or groups from anywhere in the world.

The deadline for nominations is Friday 28th September. Click here for further information.

The Strix panel hope to announce the winner at the 2018 Strix Annual Memorial Lecture in London on the afternoon of Friday, 23rd November.

To celebrate the Award and the man who inspired it, UKeiG is reproducing in this issue of eLucidate John Burchall's essay from the original Tony Kent Strix Award booklet, republished every year as the roll call of luminaries it celebrates expands. Previous winners are listed here.

The 1980s and the British Library SIR Project

John Burchall

In this [article], I want, in part, to look back some thirty years of information research but mainly to focus on quite a short period of time and on one particular project. This is the story of one project – the SIR project – the Schools Information Retrieval project. It may surprise many people that Tony was involved with a schools project, but he made a unique and important contribution, as I will try to explain. By one of those strange co-incidences, a few days ago I caught a glimpse of one of the other major players in that project – Jean Beck, who at that time (1980-81) was a school librarian of a boys' comprehensive school in West London.

It was her drive and enthusiasm that brought that school as one of the founders within the SIR project. At that time most schools had no computers at all but she was one of a number of people who had seen that they were going be an educational trend of great significance and so she brought a great pioneering spirit into that project. In fact, Jean went on to play a much wider role in bringing in new technology into school libraries and educational material generally, and is currently one of the senior staff on the National Council of Educational Technology. But first I want to put the SIR project in context.

British Library funded research

Modern information and library research goes back some thirty years, if you consider that it began with any seriousness in the work that was done at Cranfield under Cyril Cleverdon. His pioneering work involved the experiments that produced ideas about indexing, relevance and recall, and provided tools for assessing the performance of information retrieval systems. It was certainly that work that alerted a wider audience to the distinct role that research has to play in understanding information, understanding information retrieval and improving information services.

Government funding for such research was then through the Department of Education and Science's Office of Scientific and Technical Information (OSTI), which in 1973 became the

British Library Research and Development Department. There were several strands to this research over the years.

One of these strands was research into information retrieval techniques, the kind of work that Tony and others were undertaking on chemical information into developing techniques that would improve information systems. Such work still continues at Sheffield and City Universities among others, to improve the efficiency and the usability of information retrieval systems, seeking to apply the latest developments in technology to information retrieval. Chemical information services have always been at the leading edge of new developments so Tony's work at Nottingham made a great contribution to the development and understanding of information retrieval services.

The second strand sponsored by the British Library was concerned with the human side of information compared with the more machine-based information production and retrieval. During the 1970s there were many studies of information needs, information provision and how people used information in various disciplines. There were, for example, studies in chemistry, in physics, life sciences, medicine and the humanities. Some of these were large studies, reviewing in broad terms the supply of information in a particular topic or area. In addition there were a number of small studies on how people, whether practitioners or end users, sought and used information. At that time the Centre for Research in User Studies was formed at Sheffield University and became a focus for this aspect of research. The Centre undertook many studies and collected information on how people actually responded to the systems they were provided with, and how they went about hunting for information.

The third strand of research was on the education side. It is one thing to have wonderful systems but people need to know how to use them effectively and that's true both for end users and for information professionals themselves. During the 1970s, university librarians in particular were grappling with the problems of how to teach people to use the information resources that were available, both the traditional printed sources and the incoming on-line services. They felt with the introduction of on-line services there was a greater need to teach students and lecturers how to use the services to obtain the best advantage. Intermediaries are one approach but they cannot always be on hand and anyway some people like to do their own information gathering. So libraries devised various ways of teaching the use of information resources and services. Alongside this, there was concern on how the professional in the information and library world could develop expertise in using information retrieval services. During the 70s there were a lot of projects, in the main small projects, done in various library schools dotted around the country exploring different ways of producing teaching materials and packages to enable students to get to grips with computerised information retrieval. All had some support from the British Library and a wide variety of such packages, simulations and other means of teaching information retrieval were developed.

They were crude by today's standards and I recall that about the time when I went into the Research and Development Department in 1978, we had invested in a Hewlett Packard

machine, a precursor of the PC, which used a tape as the main storage device. I have a feeling that it cost around £2,000 and did very little compared with what one can get for that price today.

Information skills in schools

By the mid 70s we in the British Library felt that the ability to find and use information was not something just for the academics, researcher or industrialists but it was something for everyone. So alongside the efforts to educate university scientists and information professionals, there was a view that everyone needed some ability in information handling, to be developed at an early age. So we began a programme of projects on teaching information skills in schools. This began with some exploratory studies, to see what was happening about teaching library and information skills in schools. For example, was anyone teaching these skills on how to use libraries and how to find information from books and encyclopaedias? We found that generally there was a huge gap plus a reluctance of teachers to enter an area that they understood little about. It was in that context that the SIR project was developed.

The SIR project

It began one afternoon at the end of the 1979 Cranfield Information Retrieval Conference when a small group of conference attendees plus some teachers met to chat over the future on what was starting to happen with the use of computers in schools. Some schools were beginning to get into computers and starting to realise that the computer had a role outside the school computer department. There was also the beginning of computer based educational approaches. That afternoon led on to a longer two-day seminar at the Library and Information Department at Loughborough University. The idea of that was to bring together a much larger number of teachers to show them the sorts of things that were happening in library schools in teaching information retrieval; and at the same time to mix them with some people from the information profession and to demonstrate some of the large-scale bibliographic and IR systems. Tony Kent was one of those who came, not with his Chemical Society hat on but more as a member of the ICSU-AB (International Council of Scientific Unions-Abstracting Board), which was a forum of the large database providers. ICSU-AB had already shown interest in undertaking some educational work in schools, and Tony, for ever looking for something new, saw it as another way forward. And in his inimitable way he enthused over computer teaching tools. The meeting enthusiastically adopted the idea of developing software in order to teach information handling techniques. Tony, never surely one to hold back, offered to write the program for the experiment.

So he became a member of a Working Group set up to specify the software for the SIR project and to test it out within schools. The Working Group started in late 1979. As these things do, it took a bit longer than hoped, as all the problems had not been appreciated. I suspect that Tony got rather impatient with the delays. One feels that in his mind he already knew exactly what he wanted to do and wanted to get on and do it, whereas the other members of the Working Group were aware of the difficulties of selecting schools, encouraging teachers to participate and selecting the most appropriate computers to use

in the project. On one occasion, discussing the timetable and worrying about bringing the tools in to the schools for them to act as guinea pigs, someone asked what would happen if the software doesn't work properly. Tony said (if I recall correctly, with a smile) 'My programs always work'. He was right, his programs always did work properly, right from the start.

By the autumn of 1980 we had developed the software plus a trial database and a set of material so that the trial schools could get going. The idea was simply to create an information retrieval package, which would look like the large commercial on-line systems and would also offer similar experience in information retrieval. It was quite a tall order given the state of computers available for schools at the time. The most popular machine was the Research Machines RML380Z together with the Commodore PET, the BBC machine was also just beginning to emerge. Many schools had no computers; those that did had machines often without disk drives, instead being fitted with tape drives for storage. Capacity was severely limited, 32K sticks in my memory as the minimum requirement in which SIR would operate. Tony wrote the software in a very short space of time, I suspect he had already thought it completely through in his mind.

SIR was a means of illustrating in a practical way the essential principles of information retrieval and dissemination of bibliographical information, and also provided the means of creating files of such information. SIR was not a toy, it was not a game, it was a genuine small information retrieval system which searched databases using standard command language and Boolean logic, provided retrieval via inverted files along with an editing and database creation set of programs. It was a remarkable achievement given the state of the technology at that time. It was miniaturisation on a grand scale, brought down to the bare essentials and yet it worked and gave a feel of what you would do if you were going to go to a major online system.

It was enthusiastically received in the six schools used as a first set of guinea pigs to try out SIR. They wholeheartedly liked it in spite of the program being quite demanding. In order to understand and execute a single term search the user had to assimilate some 26 concepts, and commands, things that children and teachers had never come across before, things like database structures, record fields, commands like FIND, SHOW, PRINT, OR, AND. It was a great teaching aid for all these concepts. Nevertheless it appealed to all ages. Initially we thought that it would appeal mainly to 6th formers, but in many schools the 6th formers were the least enthusiastic. It was often students lower down in the age range who latched on to the program and showed considerable ability in grasping these concepts and using them. In spite of often only having one machine in the whole school that did not give them much chance of hands-on experience, many did begin to understand what the system could offer and to use it effectively.

The impact of SIR

SIR was a great success and went on to be marketed by Research Machines Ltd and was rewritten for the BBC micro. At that time schools were desperate for good software to use

on the computers they were beginning to acquire. SIR offered something that was immediately useable, that people could latch onto and to which children responded. It achieved what it set out to achieve. So much was owed to the fact that Tony could design programs, write them and they worked. There was considerable interest from other countries. It was even financially successful as the British Library received a royalty from the sales of the program and recouped a fair amount of the development money.

SIR helped to get good information practice into schools and the school curriculum; it helped school librarians and school resource managers redefine their role and indicated that they had a central part to play in the IT area. It led on to the use of other techniques such as CD-ROMs, specially designed networks for schools and latterly onto Internet use in schools. The fact that the SIR project had gone into schools and was in the educational mainstream meant that people were beginning to be alerted to the importance of information and the value of retrieval by computer. I would like also to think that it helped in the democratisation of information and knowledge so that it becomes available to everyone. And that everyone will have the necessary skills to find and use information effectively. There is still a long way to go, but I believe that SIR and Tony's key contribution to that project has helped in achieving that goal.

Notes for Contributors

eLucidate is the journal of the UK electronic information Group. It is published three issues per volume. It aims to keep members up to date with developments and innovations in the digital information industry, considering the impact on information professionals and consumers of e-information.

UKeiG encourages the submission of articles, reports and reviews about any of the topics covered by the journal. These include: electronic resource awareness, information management, digital/information literacy, effective information retrieval and search technologies, intranets, social media, open access, e-publishing and e-industry research and development. UKeiG can't pay contributors, but you will retain your copyright and will be able to republish your work elsewhere.

Please follow these simple guidelines:

About our members
Our membership is eclectic and includes information professionals at all levels of the UK workforce involved in digital content management and awareness, information dissemination, training and service delivery. The UKeiG demographic comprises academia, but also the private, commercial and public sectors, embracing schools, further and higher education, the NHS, healthcare and pharmaceutical industries, science, law, finance, arts, humanities, archives, museums and libraries.

UKeiG's most popular CPD courses include search tools and strategies, knowledge management, open access and research data management.

A key benefit of membership is that the CPD courses, meetings and networking forums provide "crossover" insight from one discipline to another. Members see UKeiG as a way of keeping up to date with trends and developments outside of their core, day-to-day business. Few other organisations provide this kind of cross-sectorial context and oversight.

Technical level
Although members rate themselves highly for technical awareness, they are typically users rather than creators of technology. Articles should not assume understanding of technical terms without explanation.

Length of article
Feature articles should be in the region of 1500-2500 words, but the editor is flexible on article length. Each article should be prefaced by a short summary (around 50 words.)

What to write
The world is your oyster in terms of suggested themes and subjects as long as they reflect the disciplines and membership base articulated above. You should never assume that readers will be entirely familiar with your topic, so anything you can do to offer definitions, explanations, examples and context would be welcome. You should always link to suggested reading and alternative resources to enable readers to explore your article further.

While the obvious focus of the group is the UK electronic information sector, the industry, by its very nature, is global and international developments should be reported when they impact on the UK landscape.

The most valuable viewpoint you can give is that of a practitioner. While UKeiG welcomes theoretical debate, we are primarily a forum where peers can share their practical experiences and understanding. So, if something worked for you, tell the readership. If something didn't, tell the readership why not.

How to submit
Please e-mail your copy to the editor, Gary Horrocks at: *info.ukeig@cilip.org.uk* Articles should be delivered in a simple Word format. Hyperlinks to alternative/suggested content/further reading should be embedded in the text. Images are welcome if they illustrate a point or clarify a statement. Please send them separately, and also place them in the Word document in the appropriate sections. They may be in gif or jpeg formats.

Rights
By submitting an article to eLucidate, authors grant UKeiG the non-exclusive right to publish the material in any format in perpetuity. However, authors retain full rights to their content and remain the copyright owner.

About you
Please provide a 10-20 word biographical summary about yourself, alongside an email address and job title.

Editorial process
Your article will be copy-edited for spelling and for sense. If there are major changes to the article we may return it to you for your comments and approval, but most articles require only light corrections before appearing in eLucidate, and do not need a further review by the author.