Elucidate

Statements and opinions expressed in this newsletter are those of the various authors and should not be taken as being endorsed by CILIP or UKeiG.

Contents

Developing *Trove*: the policy and technical challenges

Warwick Cathro & Susan Collier

In September 2008 the National Library of Australia embarked on a project to develop a powerful new discovery service to expose the wealth of information in Australian collections. The new service, branded "Trove", was released in December 2009. Trove is not only replacing eight legacy services, but is improving the discovery experience for the Australian public and researchers by including more content and by allowing users to engage with the content. This paper will describe the policy and technical challenges which were faced by the Library during this project, and will outline the Library's plans for the further development of Trove.

Background

One of the established activities of the National Library of Australia is its program of aggregating national metadata. This activity can be traced back to the establishment of the *Australian Bibliographic Network* in 1981, and in fact further than that – to the card-based union catalogues that were built in the 1960s. The underlying purpose of this program is to assist users to locate useful items in libraries other than their own – or (thinking beyond libraries) to facilitate the discovery and location of the wealth of information resources in the nation's collections.

Prior to the mid-1990s, these metadata aggregations were used by librarians as intermediaries for the user. The advent of the Web made it easier for these databases to be used directly by end users. In 1997, the Library released the *Register of Australian Archives and Manuscripts*, an online replacement for the printed *Guide to collections of manuscripts relating to Australia*. This was the Library's first "national online discovery service", a free service aimed at assisting researchers and the general public to discover information resources held in Australian collections.

It was evident that a combination of collection digitisation and metadata aggregation could deliver even more powerful services for users. Accordingly, in 1999 the Library released

Picture Australia as a service for discovering and accessing pictures that are digitised by Australian collecting institutions. After 10 years, that service provides access to over 1.7 million images from more than 100 separate collections. Those data contributors include museums, archives and other non-library institutions, an outcome facilitated by the use of a simple and generic metadata standard – Dublin Core.

Other national discovery services followed: *Music Australia* and *Australia Dancing* (2004), the search service for the PANDORA Web archive (2005), and the *ARROW Discovery Service* (2005), now called *Australian Research Online*, which provides a single entry point for searching Australia's open research repositories.

In 2006, the Library realised a long-held goal by opening up the Australian National Bibliographic Database to the general public, through the *Libraries Australia* free search service. And in 2008 the Library launched its online service for searching and accessing digitised Australian historic newspapers. Thus, by late 2008 the Library was operating eight national discovery services.

During the period 2000 to 2006, most of these discovery services were built on a software platform called TeraText, supplied by the company SAIC, and formerly developed by RMIT University under the name SIM (National Library of Australia, 2002).

The integration imperative

By 2006 it was evident that this multiplicity of discovery services was not providing the best possible service for the user, and was hindering the Library's ability to maintain and improve the underlying software. In that year the Library undertook a review of its IT architecture, which reported in March 2007 (National Library of Australia, 2007). One recommendation of the review was that the Library adopt a "Single Business approach" to its digital library services, implying the development of a "single data corpus" which could be deployed in a range of contexts. The outcomes would be improved service for users (less aggregations to search) and more scope for the Library to maintain and innovate (by not having to manage many silo applications).

After a further planning process during 2007-08, the Library commenced this integration project in September 2008. The internal project name was "Single Business Discovery Project". The aim of the project was to develop a discovery service that would:

• Assist users by providing just one aggregation to search;

- Reflect an emphasis on information resources held in Australian collections, without being entirely limited to such resources;

- Provide access to a greater range of resources, including more full-text content that would be immediately available to users;

- Enhance ease of discovery by providing features such as improved relevance ranking and search refinement;

- Engage with users through content and metadata annotation services; and

- Reduce the number of services requiring maintenance by the library.

Late in the project this new service was given the name "Trove", meaning a "treasure trove", defined in one dictionary as a "collection of valuable or delightful things". The name derives from the French "trouver", a verb meaning to find, or to discover. The name thus suggests the three concepts of a collection, of treasured or valuable collection items, and the process of discovery.

The Library decided to undertake this project as an in-house development, rather than use a vendor's product. In the Library's experience, vendor products provide satisfactory support for mainstream library functions, but the Library's preference is to use its own IT staff to develop more innovative digital library systems, usually based on appropriate open-source software platforms. In this case, the in-house approach was further warranted by the unusual nature of the Trove data store, including digitised newspaper and Web archive content.

Collection views

In the lead-up to the project, there was much discussion about whether users could be provided with a single relevance-ranked result set covering all types of information resource. The project team concluded that it would be impracticable to deliver meaningful relevance ranking within such a single result set because of the different resource types and metadata schemas involved. Thus, Trove has been designed to deliver multiple result sets or "collection views".

These "collection views" recognise the unique requirements of each type of information resource. Each view has its own home page, its own relevance ranking algorithm and its own facets, influenced by the type of material included in the view. Each view will define a set of "external targets" which will present additional results in those cases where important content or metadata could not be included in Trove's own data store.

Many of these views are based on the format of the information resource (for example, "Pictures and Photos"). Others reflect a topic approach (for example "About People and Organisations"). The number of collection views, and the boundaries between them, may change as the scope of the service develops, and in response to feedback from users. In the future, there may be additional views reflecting disciplines (such as "Music") which are represented in multiple resource formats, or views reflecting a wider contextual perspective (such as "Research").

At an early point, the project team conducted a "card sorting" exercise with a focus group of potential end users. The exercise was aimed at giving the team a better understanding of the format categories that were understood by users, and of the terminology which would make most sense to users in describing those categories. The report of this exercise was published on the "Library Labs" site (National Library of Australia, 2008). The exercise was helpful in answering questions such as:

- Do end users think of "books", "pictures" and maps" as format categories?

- How do end users think of "academic papers?"

- Do end users think of "archived websites" as a format category?

- Do end users often group audio and video together?

- What do users understand by terms such as "archive" and "manuscript"?

The prototype

Following this exercise, the project team developed a prototype of Trove based on the following eight collection views:

- Books, Journals, Magazines, Articles…

- Pictures and Photos

- Australian Newspapers (1803-1954)

- Music, Sound and Video

- Maps

- Archived Websites (1996 – now)

- Diaries, Letters, Archives ...

- About People and Organisations.

The prototype was released in May 2009. It included a prominent "feedback" box, and in the following six months more than 600 comments and suggestions were received. Many of these led to improvements, primarily to the interface design, and bug fixes. The prototype attracted

some international recognition. One prominent commentator described it as "one of the best one-stop shopping discovery portals I've seen" (Tennant, 2009).

Content and coverage

The wide coverage of Trove has been assisted by the Library's activities in areas such as:

- Managing Libraries Australia;
- Building the PANDORA Web archive;
- Previous involvement in the ARROW Project; and
- Undertaking the digitisation of Australian newspapers.

The coverage can be illustrated by the following concrete example. Suppose that a scholar is researching the life and works of Ethel Turner, the author of "Seven little Australians". Through a single search of Trove that scholar would be able to access the following information which would be presented on a single Web page, grouped into several result sets:

- Books by and about Ethel Turner, with information on the location of those books in Australian libraries, and with access to the full content where the work is out of copy-right;

- Articles, conference papers, theses and other research dealing with Ethel Turner, including content from university open access repositories and articles in e-journals (and in the latter case, with support for navigation to the full content where the scholar's library has a subscription to a product containing the e-journal);

- Pictures of Ethel Turner from libraries, museums and archives, including digitised pictures and information about the location of pictures not yet digitised;

- Newspaper articles dealing with Ethel Turner, and published prior to 1955;

- Archived Web sites that refer to Ethel Turner;

- Music, sound and video resources, including audio books and information about the ABC television series of *Seven little Australians*;

- Information about papers, letters, diaries and other records relating to Ethel Turner that are in archival collections; and

- Biographies of Ethel Turner from sources such as the Australian Women's Register, the Dictionary of Australian Biography Online, and Wikipedia.

Structured metadata sources used by Trove include:

- Bibliographic and authority records from the Australian National Bibliographic Database (ANBD);

- Dublin Core metadata from Australian Research Online (ARO) and Picture Australia (existing National Library discovery services);

- Dublin Core metadata from OAIster, a union catalogue for worldwide university and other academic repositories, maintained by the University of Michigan until October 2009, and thereafter by OCLC;

- Metadata from the Open Library (http://openlibrary.org/), an Internet Archive project which aims to create a Web page for every book in the world – Trove harvests only those records that link to full text;

- Metadata from the Hathi Trust (http://www.hathitrust.org/), a digital archive of library materials converted from print that is co-owned and managed by a number of academic institutions – Trove harvests records only for works that are in the public domain; and

- Tags from Wikipedia (where an ISBN is found in Wikipedia, the title of the Wikipedia article is used as a tag).

Trove also includes full text in which every word is indexed. The key sources of full text include:

- Manuscript finding aids – currently only those from the National Library;

- A selection of 120,000 books from the Internet Archive;

- About 1400 e-books from the University of Adelaide (http://ebooks.adelaide.edu.au/) (a collection of classic works of literature, philosophy, science, and history); and

- Descriptions, sample chapters and tables of contents for 400,000 books, harvested from the Library of Congress.

Biographical data

The inclusion of biographical data in Trove reflects the fact that the Library has been building a data contribution program called "People Australia" (Dewhurst, 2008). This program aims to construct a virtual Web page for each person or organisation in this data aggregation, and to support navigation to biographical information contained in other web-based services.

In parallel with the development of Trove, the project team developed software to:

- Harvest the biographical data and to convert it into a schema compatible with the Encoded Archival Context (EAC) schema (http://eac.staatsbibliothek-berlin.de/);

- Support the matching of harvested names with those already in the database, and support action by reviewers to resolve uncertain matches; and

- Expose the biographical data through standard machine-to-machine protocols (OAI, SRU, opensearch).

By September 2009, the following biographical data had been harvested:

- The Australian Name Authority File;
- The biographical data in Music Australia and Australia Dancing; and
- The Australian Women's Register.

Many other potential contributors have been identified, including the Australian Dictionary of Biography Online, the Dictionary of Australian Artists Online, and Bright Sparcs (which documents Australian scientists). At the time of writing, the National Library was in discussion with the Australian National Data Service (ANDS) concerning a possible project to extend the People Australia program by harvesting information about Australian university researchers.

This program is raising some data issues, given the sparse information in the Name Authority File and the existence of some duplicates in that file. In addition, many important Australians are excluded from the data harvested so far. For this reason, Trove will include extracts from biographical entries in Wikipedia, with links to the full Wikipedia entries. Wikipedia is considered to be a source of reasonable data quality and, although less authoritative than (say) the Australian Dictionary of Biography Online, it is a comprehensive and up-to-date source suitable for supplementing the harvested biographies.

Technical issues and challenges

Software

Trove was developed in-house by the Library, using the Java programming language. Trove uses Lucene for indexing and searching both metadata and full text, and Solr, a Web service that makes commonly used features of Lucene easily available. The project chose these tools because they are open source, and because Lucene is fast, flexible, reliable and scales very well. The Library had already used these tools successfully in other projects (the development of Australian Newspapers, and the renovation of Australian Research Online). These tools have a large and active developer community.

Trove also uses:

- MySQL as the central metadata store (http://www.mysql.com/);
- Restlets as the service framework (http://www.restlet.org/);
- Jetty as the HTTP container (http://jetty.codehaus.org/jetty/); and
- FreeMarker as the templating language (http://freemarker.org/).

Architecture and data harvesting

Trove uses four Lucene indexes: a main index, and separate indexes for the PANDORA Web archive, the newspaper content, and the People Australia content.

Currently Trove is relying on two OAI-compliant data harvesters. One of these, the Picture Australia harvester, is based on the MPS software from SAIC (http://www.teratext.com/products/teratext-metadata-publishing-system.asp). The other ("NLA Harvester") was written by the Library, and is being used to harvest the data for the Australian Research Online and People Australia programs. Once the NLA Harvester has been extended, it will replace the Picture Australia harvester.

Currently Trove does not have enough capacity to index every item of full text that it could harvest from the sources described above. At the time of writing, the equivalent of about 120,000 full-text books had been indexed.

Update performance

One downside of Solr is that it is relatively poor at supporting systems where record updates need to be quickly visible. This has been a challenge, because it is desirable for Trove to incorporate changes such as user tags and comments in near real time. As at January 2010 it can take up to one minute between making a change such as adding a tag and having it visible in the interface. The project has judged this delay to be acceptable at present, and by mid-2010 hopes to benefit from the real-time update work being done as part of the Lucene project (Apache Lucene, 2009).

Grouping of Works and Versions

It will be recalled that one aim of the project was to make discovery "as easy as possible" for the user. Part of this involves making the presentation of result sets as intelligible as possible for the user. A framework supporting this aim is FRBR (Functional Requirements for Bibliographic Records) which uses the hierarchy Work, Expression, Manifestation, Item (IFLA, 2009).

The Trove result set display does not implement the full FRBR model, but focuses on grouping records into Work and Version clusters. A Version can be considered as a level intermediate between Expression and Manifestation.

Trove attempts to group each record from any of its contributing sources with other records in the database. When a user searches the system, each grouped record appears as a single search result. The rules used to group records into a Work cluster are as follows:

- A record is classified into a format such as book, picture, or map. Only records of the same format type can be grouped together.

- A scoring system is used to determine whether two records are a match. Records are assigned a score for each match point found. If a threshold score is reached, the records are grouped together. Match points include:
  - OCLC Work Identifier (this is considered a definitive match)
  - Authors and Creators
  - Title, Uniform title, Other titles
  - ISBN, ISSN, LC Control Number
  - Publisher (very low scoring).

For example, "Alice in Wonderland" the talking book, by Lewis Carroll, would be grouped in the same Work cluster as an ordinary text version of the same book, but not with a video of "Alice in Wonderland". The edition of "Alice in Wonderland" by Lewis Carroll, published in London in 1966 would be grouped in the same Work cluster as the edition published in America in 1982.

A Work cluster is further sub-sorted into Version clusters. Language, sub-format, publication dates, edition statement and/or numbers, publication details, physical description, author and series are used to assign a record to a Version cluster. If all these fields match in more than one record, then the records are grouped into a single Version cluster. Each Version cluster will potentially record multiple "Items" held by multiple libraries.

Since these matching algorithms can never be perfect, users have been given the ability to amend the Work and Version clusters. A user can merge one work (Work A) into another (Work B), in which case Work A becomes a Version within Work B. Conversely, a Version can be separated from a work to become its own work. The user is able to add a note to each change to explain their reasoning for the move. To date a small number of interested users have used this feature.

Online access

It is important that Trove allows users to limit their searches to "online resources", but it has proved difficult to determine what level of online access is provided from a URL in the

metadata record. In some cases this URL links only to a table of contents or a publisher's description, and in other cases a subscription is needed to enable access to the full resource. The project team undertook extensive analysis to identify clues to the level of access. Trove supports three categories of link: "available online", "available on line (access conditions apply)", and "possibly available online".

User interaction

Australian Newspapers, released as a beta service in July 2008, provides for significant interaction by the user community, including the ability to correct the OCR text, and to add tags and comments. Trove will extend this kind of user interaction to all collection views. Users will be able to add tags or comments to any record, and this user-contributed data will be searchable in Trove. In addition, as we have seen, users will be enabled to move a record from one collection view to another, and to amend the work and edition groupings.

Future developments

The National Library has set out its plans for the future of Trove in the Strategic Plan for the Resource Sharing and Innovation Division (National Library of Australia, 2009).

The Library implemented the first stage of Trove during December 2009, accompanied by the decommissioning of two legacy software applications as stand-alone services: the Libraries Australia free search service, and the Register of Australian Archives and Manuscripts. Another stand-alone service, the Australian Newspapers service, will be decommissioned in the first quarter of 2010.

The Library is now working on further improvements to Trove. As at January 2010, it is making a number of general improvements such as provision for users to create lists for a range of purposes, RSS feeds, enhanced sorting of results, more external targets, indexing of more full text, further improvements to the user interface, and bug fixes. The Library then plans to develop the Australian Newspapers collection view to meet priority enhancements identified during the Australian Newspapers Beta phase.

During 2010, the Library plans to commence the decommissioning of other legacy services, commencing with Picture Australia. This decommissioning will be actioned only when Trove has at least matched the user interface and functional features of the legacy service, including the provision of trails and advanced searching.

In the second half of 2010, the Library plans to expand the journal article content in Trove. This will involve:

- Support for search and delivery of the full content of any journals digitised by the Library;

- Inclusion of journal article indexing data and the linking of this data to library holdings; and

- The ability for Australian library users to easily discover and link to e-resources, from selected vendors, which they are entitled to access by virtue of their library member-ships.

The Library aims to strengthen the coverage of Trove by exposing content held by archives and museums. The preferred strategy is to obtain high level collection guides, finding aids and file descriptions, to index the text of these guides, and to rank them highly in the "Diaries, Letters, Archives" search results. The user will navigate from the copy of the guide in Trove to the original guide at the archive or museum Web site and, where possible, to specific items within the collection, where these are documented on that Web site. This strategy will effectively provide a context for the user and an appropriate way of linking to archive and museum collection content.

Conclusions

The release of Trove follows three years of discussion, analysis and development by the National Library, a process which began with an IT Architecture review in 2006. Trove has taken the Library's discovery services to a new level, in terms of the degree of integration, the expansion of full text content, and the extension of user contribution to content categories beyond newspapers.

Many integrated portals have been developed by libraries, library networks and vendors. Trove is unusual in its breadth and in its national scale, given its inclusion of archived Web sites, digitised newspaper articles, university research outputs, the national union catalogue and online biographical databases.

References

Apache Lucene – overview (2009). http://lucene.apache.org/java/docs/index.html

Dewhurst, Basil (2008). People Australia: a topic-based approach to resource discovery.
http://www.valaconf.org.au/vala2008/papers2008/116_Dewhurst_Final.pdf

IFLA (2009). Functional requirements for bibliographic records.
http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records

National Library of Australia (2002). Media Release. 12 August 2002.
http://www.nla.gov.au/pressrel/2002/teratext.html

National Library of Australia (2007). IT architecture project report
http://www.nla.gov.au/dsp/documents/itag.pdf

National Library of Australia (2008). Single Business card sort report, December 2008. Linked from https://wiki.nla.gov.au/display/LABS/1.+Our+Prototypes

National Library of Australia (2009). Resource Sharing and Innovation Division. Strategic Plan, July 2009 to June 2011. http://www.nla.gov.au/librariesaustralia/documents/div3-strategic-plan.pdf

Tennant, Roy. One-stop searching with a can-do attitude. 26 May 2009.
http://www.libraryjournal.com/blog/1090000309/post/1900044990.html

Acknowledgments

Warwick Cathro (wcathro@nla.gov.au) is Assistant Director-General, Resource Sharing and Innovation at the National Library of Australia, and Susan Collier (scollier@nla.gov.au) is Project Manager for the Trove Project.

Mining information-seeking behaviour data to enhance library services

Jenny Walker

Sitting on a Gold Mine

'Google Wants Your Links, Not Your Content' is the title of a recent posting to the Society for Scholarly Publishing (SSP) blog, which draws our attention beyond the content to the use of that content. [1]

Users indicate their preferences and selections as they navigate between sets of resources and this valuable data—user 'clickstreams'—can be collected and analysed. User click-streams, indicating the actual selection of users, are being used by service providers such as Google[2] and Amazon[3] to help users discover items of interest. To date, user information-seeking behaviour data has largely been overlooked for enhancing library services, but steps are being taken in this direction, such as the development of recommender services and new metrics for scholarly evaluation.

If you looked at this item, you might also be interested in ...

Recommender services are now appearing in library applications. LibraryThing[4], BibTip[5] and bX [6] are such examples. These applications record the information resources users access and the order in which they access them, in order to make suggestions for other users. LibraryThing and BibTip record access to the 'catalog' and recommendations are offered at the title level, typically the book. bX can potentially record access to the whole library collection, including remotely-hosted resources; recommendations are typically offered at the article level.

BibTip and bX, deriving from research projects at Karlsruhe University, Germany and the Los Alamos National Laboratory respectively, use a statistical analysis of user information-seeking behaviour to generate recommendations. Recommendations are offered based on co-retrieval of items within a user's session. BibTip uses data from local OPAC usage, while bX aggregates link resolver usage logs from multiple institutions around the world.

BibTip and bX are good examples of harnessing collective intelligence from library users to serve the needs of the library. Recommender services proactively help the user find

information without requiring explicit user queries; interesting items find the user instead of the user explicitly searching for them.

Metrics for scholarly evaluation

The harnessing of collective intelligence from library users is also being explored for the provision of new metrics for scholarly evaluation. Although in the last decade the scope of scholarly communication has broadened well beyond the print environment, the evaluation of research is still largely based on citation and authorship data and has its genesis in the print domain.

User-driven evaluation offers an interesting alternative to citation-based evaluation; shifting the focus from authorship to readership, this alternative offers more immediacy in reflecting the importance of articles for users, and could be especially helpful for journals with high undergraduate or practitioner use. Further, it has the potential to cover new materials and new types of material not currently covered by the Journal Impact Factor[7]. Metrics based on usage are unlikely to replace the well-established impact factor, but could be an important complement.

There are a number of current initiatives towards the determination of usage-based metrics for scholarly evaluation, including the United Kingdom Serials Group (UKSG)[10] Usage Factors project[8] and project MESUR[9].

UKSG Usage Factors

In 2006 UKSG commissioned a project to investigate the potential for usage data as a way of generating metrics for scholarly evaluation. The starting point was the vast collection of COUNTER[11]-compliant usage data.

Positive indications emerged from the results of surveys conducted with librarians and publishers (2006-2007); and from the subsequent testing and modelling with real usage data (2008). The next steps to be undertaken aim at identifying potential candidate usage metrics for longer-term testing on a grand scale. These involve data analysis and modelling using data from a number of content providers.

Project MESUR

Project MESUR, lead by Johan Bollen and Herbert Van de Sompel from the Los Alamos National Laboratory, USA, and supported by the Andrew W Mellon foundation, has earlier this year reported on the outcome of their investigations into usage-based metrics.[12]

The MESUR team collected more than a billion transactions from OpenURL link resolvers and significant scientific publishers and aggregators. These transactions reflect user behaviour across a wide and diverse set of scholarly resources, and represent electronic data searches in which users moved from one journal to another, thus establishing associations between them.

Project MESUR has surveyed a number of different citation- and usage-based metrics (nearly 40) that each represent a unique perspective on scientific impact. Some key dimensions emerge along which scientific impact can vary: most particularly the speed with which a metric indicates changes in scientific interests over time, and also the popularity of a journal versus its prestige or influence. As such each metric expresses a mix of these aspects of scientific impact, and can be selected to favor one or the other.

Map of Science

In addition to proposing usage-based metrics for scholarly evaluation, the MESUR team used their large set of usage data to create a detailed and contemporary view of scientific activity (fig.1). Each dot on the map represents a journal, and the journals are colour-coded for easy subject recognition. The interconnecting lines reflect the probability that a reader will move from one journal to another on the computer screen, each time clicking on articles of interest.

Figure 1: Map of science

This map differs significantly from similar maps constructed on the basis of citations rather than usage, and corrects the under-representation of the social sciences and humanities that is commonly found in citation data. According to Dr Bollen, clickstream maps offer an immediate perspective on what users are doing, and can therefore assist in the detection of emerging trends, inform funding agencies and aid researchers in exploring interdisciplinary relationships.[13] Further, such maps can help researchers to identify important journals in any particular domain of interest.

Next steps

These first steps in mining user behaviour data to enhance library services are important ones and set libraries on the road to appreciating the value locked up in the data they hold. Information-seeking behavior patterns can serve for a better understanding of the links between items that make up a library, enable better guidance in the use of library resources and can help assess the value of scholarly materials. With our society increasingly focused on measuring research outputs and research quality, serious consideration must surely be given

to new usage-based metrics. In the future, combining user clickstreams with user-profile information has the potential to make this data even more valuable.

References

1. Anderson, K, Google wants your links, not your content. April 14, 2009. Available at: http://scholarlykitchen.sspnet.org/2009/04/14/links-matter-more-than-content-folks/
2. Google
   http://www.google.com
3. Amazon
   http://www.amazon.com
4. LibraryThing
   http://www.librarything.com/
5. BibTip
   http://www.bibtip.org/
6. bX Recommender Service
   http://www.exlibrisgroup.com/category/bXOverview
7. Thomson Scientific Journal Impact Factor
   http://thomsonreuters.com/business_units/scientific/free/essays/impactfactor/
8. UKSG Usage Factors
   http://www.uksg.org/usagefactors
9. Project MESUR
   www.mesur.org
10. UKSG
    http://www.uksg.org/
11. COUNTER
    http://www.projectcounter.org
12. Bollen J, Van de Sompel H, Hagberg A, Chute R. 2009. A principal component of 39 scientific impact measures
    Available at: http://arxiv.org/abs/0902.2183
13. Bollen J, Van de Sompel H, Hagberg A, Bettencourt L, Chute R, et al. 2009 Clickstream Data Yields High-Resolution Maps of Science. *PLoS ONE* 4(3): e4803. doi:10.1371/journal.pone.0004803. Available at http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0004803

Note: this article first appeared in Serials, the UKSG newsletter, Volume 22, Issue 2, pp182-184.

Jenny Walker is an Information Industry Consultant. Prior to starting her consultancy practice in 2008 Jenny held a number of senior marketing roles with technology and content providers, including executive vice-president marketing at Credo Reference, vice-president marketing at Ex Libris, and director of technology product management at SilverPlatter. Jenny has a keen interest in the development and deployment of interoperability standards and currently serves on the NISO architecture committee.
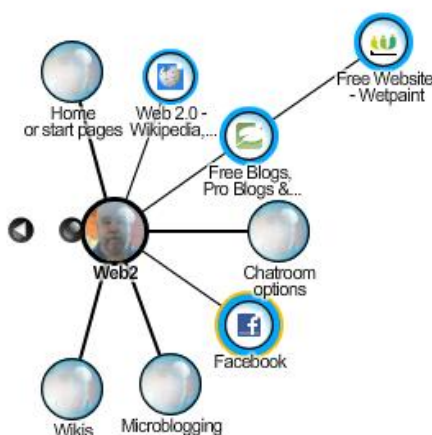
Web 2.0

Phil Bradley

This month's column looks at some recent Web 2.0 products

Some recent Web 2.0 resources

I'm very fortunate in that I'm able to spend a lot of my time exploring new resources to see what they do, how they work, and how they compare with their competition. In this column I thought I'd look at a few of the new products that have passed by my monitor in the last few months. No particular order this time around, which is actually part of the fun that I have – I'm never sure exactly what I'm going to find next.

Pearltrees

One of the things that information professionals do well is to seek out good quality resources, which can then be passed onto colleagues. However, while the simple emailed list does the job perfectly well, there are a range of what I call 'guiding tools' which are now entering the marketplace. One of these is Pearltrees at http://www.pearltrees.com/, which allows users to create 'pearls' or trees of websites in a visual fashion, similar to mind mapping, and indeed the resource could be used for exactly that as well. URLs can be linked together in a typical tree style with subject branches as appropriate. I did a very brief one which you can find at http://pear.ly/ngyL which should be enough to give you a good idea as to how they work.

It's very easy to add new 'pearls' to an existing Pearltree by means of their handy little bookmarklet, so when you find a resource that you want to add to a tree it's the matter of a

few moments to add it as appropriate. Pearltrees also emails you if someone else uses the same URL(s) in their trees, so it's easy to keep tabs on who else is covering the same subject matter. It's then possible to jump from one persons tree to another, via the same connection. Alternatively, search Pearltrees by subject to see if anyone has created a tree in a subject area that interests you.

Tagxedo

This is a word cloud generator, and you can find it at http://www.tagxedo.com. It's the latest in a long line, with perhaps the most well known being Wordle at http://www.wordle.net. For several years Wordle has ruled the roost when it comes to visualisation tools – just provide it with an RSS feed (such as a weblog) or paste in some content, and it will then produce a cloud of words – the larger the word, the more often it's mentioned. You can change the colours, font size and direction of words, and it's a fun tool. However, even good tools, if they're not updated with new functionality, fall victim of something better. Tagxedo is the next generation word-cloud generator, since it does everything that Wordle can, but more besides. Users can upload their own images and Tagxedo will create a word cloud around that shape. It's particularly good for silhouettes (I created a nice Sherlock Holmes for example) and by

cutting and pasting in the content from a Wikipedia entry for example you can get some striking images.

[The embed code for my image is:] Moreover, as Tagxedo uses Microsoft Silverlight technology (which you may have to download before you can play with it properly) if you mouse over particular words they jump out to the front, so it's a very useful visual aid or training tool.

WebKlipper

Another way of providing content quickly and easily is to use a resource such as WebKlipper at http://webklipper.com/ The idea behind it is very simple – simply give it the URL of a page that interests you, or which you wish to comment on and WebKlipper loads the page into a frame onto their website, where you can highlight text, comment on it, extract and share it. WebKlipper then saves the page for you, together with your comments in its own URL that you can direct friends or colleagues to. You can also subscribe to a particular page and see if anyone else comments as well. Moreover, you can create your own page, for free, and without any signups either. It becomes a very simple blog page or website page if you want to share something quickly and easily.

Checking website status

There's nothing more frustrating than trying to reach a website only to get some message or other which essentially says 'tough, this site isn't playing.' There's always a niggle at the back of your mind as to whether the site IS actually live, and it's your connection that's down. I've found three resources that will check to see if a site is up and running, and these can at least set your mind at rest, enabling you to go about the rest of your business knowing that no-one else can access that important site either. The first of these is the aptly named Downforevery-oneorjustme.com at http://www.downforeveryoneorjustme.com/, which I'll grant is a long URL, but it's also one that you're unlikely to forget though as you'll read in a moment, not necessarily the most memorable. Simply type in the URL that you want to check and it will tell you (generally within a second) if the site is, well, down for everyone or just you! The second site, which works in exactly the same way, is Netirk at http://www.netirk.com/ and it works just as quickly. The last of the three is the rather laddishly named Titsup which has the interface 'Has _____ gone tits up?' and you can find it at http://www.titsup.net (A word of warning here – other variants of the domain name such as .com will in all likelihood not be

safe for work, so you're warned in advance.) A real belt and braces approach would be to try a couple of these just to be on the safe side, since it's theoretically possible that the site you're interested in could be down for both you AND the resource you use to check, but still up for other people.

Screencasting software

I was taken via a Twitter link to a YouTube page recently, on which someone was giving a guided tour around a product. I was able to see what his cursor was doing, what he was typing into search boxes, zooming into parts of the screen and so on. Of most interest however was the fact that there was a small webcam image of him in the corner of the screen and this gave the tutorial an entirely different dimension – much more personal and informative somehow. Of course, once I'd seen it, I knew that I had to have a go at doing something like that myself! After a bit of trial and error I found a package called BB Flashback Express at http://bit.ly/9a0YuE which is a free product (registration required after 30 days however) although there is a commercial version available. It does require a download (if you're not allowed to download and install software, try Screentoaster at http://www.screentoaster.com which is browser driven) but is very straightforward. You can then use it to record your screen, or part of your screen and incorporate your webcam, or if you prefer, just an audio commentary, and then save the movies as Flash or AVI, installing them online or uploading them to YouTube, Blip.tv, Viddler or Revver. It's compatible with XP, Vista and Windows 7. You can use it to record something in a browser window, but that's not necessary, so if you wanted to create an online guide to using a library catalogue for example, it would be simplicity itself. If you want to take a look at a worked example I did a screencast of Pearltrees, and you can find it in my YouTube channel at http://www.youtube.com/watch?v=Yd091_x0WNE [If you want to embed it on the page, the code is <object width="480" height="385"><param name="movie" value="http://www.youtube.com/v/Yd091_x0WNE&hl=en_GB&fs=1&"></param><param name="allowFullScreen" value="true"></param><param name="allowscriptaccess" value="always"></param><embed src="http://www.youtube.com/v/Yd091_x0WNE&hl=en_GB&fs=1&" type="application/x-shockwave-flash" allowscriptaccess="always" allowfullscreen="true" width="480" height="385"></embed></object>]

This rather neatly brings me back to my starting point, which is as good a point as any on which to end this column!

Online

Joy Cadwallader, Aberystwyth University (Aberystwyth Online User Group)

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

Anne Frank House

http://www.annefrank.org/secretannex/

As part of the 50th Anniversary of the Anne Frank House Museum (03/05/10) a virtual tour of the Anne Frank's hiding place has been launched. It begins with the secret entrance behind a bookcase. The tour was developed by LBi Lost Boys international (http://www.lbi.co.uk/) in collaboration with the Anne Frank House. Funding was provided by the BankGiro Lottery, the Anne Frank Fund, the VSB Fund, the Dutch Ministry of Health, Welfare and Sport and the Mondriaan Foundation.

Digital Economy Act 2010

http://www.opsi.gov.uk/

The Digital Economy Bill was rushed into law late on the night of Wednesday 7th April as part of the "wash-up" normally intended to complete non-controversial business prior to a general election. Clause 17 (see Online January 2010) was replaced by a re-drafted Clause 18, which was subsequently dropped at the last minute, with a hasty amendment to Clause 8, which gives powers to the Secretary of State for Business to order the blocking of "a location on the Internet that the court is satisfied has been, is being or is likely to be used for or in connection with an activity that infringes copyright". Clause 43, permitting re-use of orphan works, was dropped following opposition from photographers, which will be a disappointment for the British Library, who continue to face extensive copyright difficulties before digitising orphan works. Clause 45 permits libraries to lend ebooks and audiobooks without infringing copyright. Ofcom has drafted a code of practice (http://www.ofcom.org.uk/consult/condocs/copyright-infringement/) as instructed by the Act in preparation for identifying and penalising persistent illegal downloaders, with a range of measures including stopping their Internet connections. ISPs will be expected to implement these measures or risk fines of up to £250,000. Ofcom has asked ISPs to start collecting data on illegal download activity (http://www.guardian.co.uk/28/05/10).

Google

The Open Book Alliance have produced a report
(http://www.openbookalliance.org/2010/05/the-gbs-makes-for-angry-neighbors/) that makes
the case that the draft Google Books settlement would violate international laws and treaties.
Intellectual property lawyer Cynthia Arato draws the conclusion that the US could be
challenged by other countries through the World Trade Organisation, with the attendant risks
of fines, trade sanctions and "diplomatic stress".

A new class action has been brought against the settlement by the American Society of
Media Photographers, together with other creative organisations and individuals, challenging
Google's right to digitise images in books without rightsholders' permission.

JSTOR

http://www.jstor.org/

JSTOR have announced their new Business III collection, starting with seven titles and
building to a minimum of 50 by the end of 2012. Three of the titles due to be included have
enjoyed over a century of a publication: *Journal of Institutional and Theoretical Economics*
(1844), *FinanzArchiv* (1884) and VSWG: *Vierteljahrschrift für Sozial- und Wirtschafts-
geschichte* (1893).

Library of Congress

http://www.loc.gov/

Accompanied by a rush of headlines: "How tweet it is!" (Library of Congress); "All your Twitter
belongs to the Library of Congress" (Guardian); "OMG! Library of Congress acquires every
Tweet ever Twittered" (LA Times); the Library of Congress announced that Twitter is donating
the entire public tweet archive since 2006, comprising billions of tweets, to the Library of
Congress. Librarian James H. Billington said, "Working with the Twitter archive will ... help the
Library extend its capability to provide stewardship for very large sets of born-digital
materials." Their FAQ (http://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/) details
what the archive will or won't include. Google have also announced their Twitter archive
search tool called Replay.

Nature Publishing Group

http://www.nature.com/

The complete Scientific American archive May 1948 through December 2005 is now available online via institutional subscription, comprising more than 650 issues and 20,000 articles.

Meanwhile Nature Publishing Group (NPG) have announced open-access options for seven more journals, making a total of 25 including all 15 NPG-owned academic journals. Authors can choose to make their articles open access by paying an article processing charge (APC).

OUP

http://www.oup.com/

Free institutional trials are available for Oxford Bibliographies Online. The press release[1] describes OBO as, "an entirely new type of online resource offering a library of discipline-focused, online guides to the essential literature in subjects across the humanities and social sciences". The first seven subject areas covered are Classics, Islamic Studies, Criminology, Social Work, Renaissance and Reformation, Philosophy and Atlantic History, and each area is designed to guide researchers through the most important resources for that subject.

Paywalls

Following the introduction of *The Times* paywall (see Online March 2010) the *Guardian* has announced their free trial of Guardian Extra, "a membership scheme for *Guardian* readers", which will cost £25 pa from September 1st. It is currently marketing a different model from *The Times*, describing Extra as providing add-on features to existing content, online events and special offers, rather than ring-fencing premium content. Bloomberg (finance), *Advertising Age* (marketing and media) and *Rolling Stone* have all announced paywalls for some of their content (http://industry.bnet.com/ 20/04/10). Meanwhile Rupert Murdoch has announced a monthly, online "professional edition" of the *Wall Street Journal* priced at $49. WSJ Pro Consumer consists of WSJ Online and 17,000 Factiva publications. ResourceShelf noticed that the product is not dissimilar to WSJ Pro Enterprise launched in November (http://www.resourceshelf.com/2010/04/08/packaging-sources-for-sale-wsj-factiva-wsj-pro/08/04/10).

---

[1] See Press Releases, below

UKSG

http://www.uksg.org/

The UK Serials Group (UKSG) and NISO (National Information Standards Organisation) have announced that Phase I of their joint project for "exploring data problems within the OpenURL supply chain" has been endorsed by the American Institute of Physics, Ex Libris, OCLC and Serials Solutions. The KBART (Knowledge Bases And Related Tools) Working Group published its first practical recommendations in January 2010, and encourages all content providers to submit a sample file for approval, so they can also join a public list of endorsing providers. Members of KBART Working Group for Phase II are listed in their press release (http://www.uksg.org/news/kbartmay10/).

Intranets

Martin White, Intranet Focus Ltd (martin.white@intranetfocus.com)

Any time, anywhere

Once upon a time, when the pilot shut down the engines of the aircraft, there was a short period of calm as passengers readied themselves for the torment of the baggage carousel. Now the silence is broken by the noise of mobile phones being switched on and the beeps of messages arriving. We live in a world where being away from email and text messages for even an hour means that we are concerned that decisions have been made without our invaluable input.

Over the last couple of years the mobile handset has changed beyond all recognition. The Apple iPhone and the Google Android are changing the mobile paradigm from messaging to information and application access. However possibly the most important recent development has had much less visibility and that is the purchase of Palm by Hewlett Packard (HP). HP are not interested in Palm hardware but in the WebOS operating system and a raft of 1500 patents on mobile information access that Palm have not had the resources to develop any further.

Going back to aircraft for a moment, one of the long-term outcomes of the recent air traffic disruption in Europe from the Icelandic volcano was that staff at all levels found themselves working from hotel rooms and airports, often without a laptop, as they were supposed to be on holiday. Their smartphone could act as an effective email access device, but most were marooned from their corporate intranet. So just tracking down staff telephone numbers on the corporate directory became a nightmare.

Without any doubt the mobile handset business, driven by consumer demand, is going to be a transforming influence not only on intranets but on Web sites as well. A good indication of the rate of development of mobile accessed applications is an article in the May 2010 edition of *Chemistry World* [1] in which Antony Williams, the inventor of the chemical structure database

---

[1]

http://www.rsc.org/chemistryworld/Issues/2010/May/MobileChemistryChemistryHandsFace.asp

ChemSpider, looks at just a few of the applications which have been developed for practising chemists in industry, universities and schools.

In short, if you do not have a mobile access strategy for your intranet, then you only have a few months left to come up with one, or condemn your colleagues to being at a major competitive disadvantage. The usual response to my enquiry about why a client does not have effective mobile access to the corporate intranet is that security is an issue. When security is getting in the way of access to business-critical information then it's time to fire and hire in the IT security department. The second most common response is that providing access to all the intranet is just too much of a content-management challenge.

I would be the first to suggest that it is not necessary to provide access to all intranet content, but to focus on delivering content that is of particular value to staff who are rewarded for being out of the building and then penalised by not having access to the intranet. Indeed a good way to start is to add a new persona to your collection that summarises the key requirements of mobile staff. This is not just a senior executive issue, but ripples all the way down to young graduates to whom the mobile handset is the way they live and wish to work.

One technology issue that does need careful consideration is mobile search, though GPS information can be used very effectively to limit search queries and results to those relevant to the precise location of the user. Rarely will a project manager in Germany be interested in projects being carried out in Brunei. One of the neatest solutions I have come across to date is Isys Anywhere from Isys-Search (www,isys-search.com) but expect other search vendors to come up with similar products by the end of this year. If they do not then you may wish to consider whether you have the right search vendor.

Recent reports from Accenture, Morgan Stanley, Deloitte and the Nielsen Norman Group all point in the same direction, and the supporting evidence will be around you every time you fly, travel on Eurostar, or have a drink in a hotel bar. There is little point in investing in putting content into an intranet that at any particular moment in time perhaps 20% of your colleagues will not have access to as they are away from their desks. Don't think for one minute that people who need access will have a secure lap-top. Apple, Google and HP are betting that they will not. Are you going to bet against them?

Reference Management

Tracy Kent, Digital Assets Programme Advisor, University of Birmingham (t.k.kent@bham.ac.uk)

Reference Software

The aim of this column is to highlight those tools that support the management of references – the main stable of the information profession. The use of such tools has grown exponentially over the last decade, and now the decisions on which one to use is based on identifying where references are obtained from (propriety software, websites), the budget (open-source verses commercial) and the resources available (for example, how good is the support).

If this is a new area to you then take a look at one of several comparison pages on the Web, including

- Adept Science comparison site at http://www.adeptscience.co.uk/products/

- Beyond my mind : comparison of free bibliographic software packages http://mahbub.wordpress.com/2007/03/04/comparison-of-free-bibliographic-managers/

- UKeiG factsheets on how to manage references and what to look for in a package http://www.ukeig.org.uk/

- Wikipedia also has a useful comparison site at http://en.wikipedia.org/wiki/Reference_management_software.

Open-source and reference software

Many of us think of Linux when we hear "open source". While Linux is a great example of the open-source process, it is a small segment of the movement. Many open-source applications are available to help manage the reference process and most applications are available for Microsoft Windows and the Apple Macintosh.

Bibus

Bibus is an open-source bibliographic and reference management software tool. As with other such tools, Bibus allows you to search, edit, and sort bibliographic records. In addition, Bibus has the ability to create custom styles for bibliographies. It works with most platforms and can be integrated with Microsoft word for displaying references. Version 1.5 was released back in October 2009. Further details at http://bibus-biblio.sourceforge.net/wiki/index.php/Main_Page

JabRef

JabRef is another, popular, open-source reference manager package which uses BibTex as its native file format that is the standard in LaTex bibliography formats. It runs on the Java virtual machine which means the program works equally well with Windows, Linux and Mac OS X. All the basic features of a reference software package are available and includes the ability to export as a specific pdf document. Further details from http://jabref.software.informer.com/.

MyReferences and Moodle

JISC, in partnership with the Open University and RefWorks (the CSA based reference software package) , has developed MyReferences. MyReferences integrates tools from RefWorks into Moodle the virtual learning environment used in many FE colleges and some Universities.

It is based on open source[1], making it free for any institution to download, tailor to their own needs and integrate into their own learning environment and is the result of the TELSTAR project. Its unique strength is the ability to integrate reference management with an online learning environment. A simple click through provides users with the ability to select appropriate sources, their preferred reference style and then cut and paste the results into their assignments. Institutions need to subscribe to RefWorks to adapt the added Moodle integration functionality using RefWorks API. However, as the software is available as Open Source, institutions or individuals could take what has been done as the basis for working with other reference management software. Further information from http://www.open.ac.uk/blogs/telstar/telstar-quick-demonstration/

The software Is available at http://code.google.com/p/telstar/, and there is the obligatory discussion group at http://groups.google.com/group/TELSTAROU, open to anyone to join.

Reference Tracker

Reference Tracker creates documents that store all the citations and references and then automatically creates Harvard, APA, MLA or Chicago/Turabian formatted reference lists to insert into your word-processed document. Usefully, it allows you to add sticky notes to references and export in a number of formats. It is a Mac product, which is also useful, and has been developed in the UK. It is perhaps worth pointing out though that the *Reference Current Page in Firefox* feature is not compatible with Firefox version 3.

---

[1] a topic covered by this year's UKeiG Annual Seminar.

http://www.malkinware.com/referencetracker/

InfoWorks

Part of the reference process is ensuring URLs are correct. A tool, Infoworks Line Checker, created by librarians, is designed to do just that, by working with flat MARC and non-MARC text files to check URLs. http://www.vyomlinks.com/download/infoworks-link-checker-2.0-125606.html

MarcXimiL is a free, standards-compliant bibliographic similarity analysis framework. It allows you to analyse at different levels, run in batch mode or through the API. Options include the order of comparisons between and within collections. or comparison amongst a wide selection such as vectorial, probabilistic, Levenshtein based or maybe just Authors and Date! The software enables you to detect plagiarism while supporting quality information management. http://marcximil.sourceforge.net/

Open-source reference tools in schools

A useful service relating to open-source referencing for schools is SchoolForge, which lists available software and additional support mechanisms (including a discussion area). It is kept very up to date making it invaluable for busy information professionals.

http://www.schoolforge.net

Apple now the market leader in reference software too?

You may have read recently that Apple has passed Microsoft in the market-value stakes. This upsurge may affect the reference software market as Apple release version 6 of Sente. Sente, as well as providing traditional reference software features as well as downloading and importing PDFs direct from publisher's Web pages, takes notes from PDFs, highlighting the text in the PDF itself, and adds references using ISBN or DOI's. There is also a powerful new tagging feature. Concentrating on how users can deal with the full text is definitely the next phase of managing references, and we await to see how Microsoft respond!

http://www.thirdstreetsoftware.com/site/introduction.html

And finally : Preserving references

Being able to find references long after they have been added to your reference software is essential. There is a useful bibliography available in this area which contains over 300 articles and books on understanding digital curation and preservation. The majority of the links are available on open access.

http://digital-scholarship.org/dcpb/dcpb.htm

Public Sector News

Jane Inman, Customers and Communications Manager, Environment and Economy, Warwickshire County Council (janeinman@warwickshire.gov.uk)

Keeping track of the changes in the machinery of government is always challenging and no more so than after a General Election with a complete change of government as we have just had in the UK.

How do we keep up to date?

www.Direct.gov.uk carries a link on its home page with the headline *Key ministerial appointments.* The appointments are given and the changes to Departments.

The Department for Children, Schools and Families has been replaced by a Department for Education and HM Treasury has a new Office for Budgetary Responsibility.

A new website is up and running at http://www.education.gov.uk for the Department for Education. The publications of the old department are under review, and the link to publications and to content of the old site (which can still be searched) carries the warnings:

> Due to a change of government, a review of all Departmental (including National Strategies) publications is taking place and as such dispatches of publications are currently on hold until further notice.

And

> A new UK Government took office on 11 May. As a result the content on this site may not reflect current Government policy. All statutory guidance and legislation published on this

site continues to reflect the current legal position unless indicated otherwise.

How do we find the old websites?

The department has changed and new content is being created but how do we find information from the old department for research or general information?

The preservation of Government websites is the responsibility of the National Archives www.nationalarchives.gov.uk. To find the Department for Children, Schools and Families, follow the quick link on the home page to the UK Government Web Archive. The content is arranged by topic, and drilling down through these the old Web site is quickly found. The simple and the advanced search were not as successful.

The archive has been developed by the Digital Preservation department of the National Archives, and it preserves all types of electronic government records; it has been operational since 2003. It is split into:

- UK Government Web Archive,

- Electronic Records Online which covers the 95% of UK government records which are 'born digital', and the

- National Digital Archive of Datasets, which provides access, as its title suggests, to data sets that have been developed to 'collect, store and analyse statistical information' and that comprise the information that has been used to formulate policy and legislation so are a valuable resource.

Open data to continue

David Cameron will continue to support moves to make public-sector data freely available as begun under the previous government. In local government the example of London, which has http://www.data.london.gov.uk, and the Government, which has published data at http://www.data.gov.uk, has now been followed by http://www.opendata.warwickshire.gov.uk and http://www.lichfielddc.gov.uk/data and http://www.eastsussex.gov.uk/about/standards/**opendata**.htm. These are just the first page of results from a search of Google, so I am sure there are others.

It is anticipated that the publication of public-sector expenditure in a standard format will be required in due course and the model used by Windsor and Maidenhead has attracted some publicity. It is clear and easy to understand; it can be found at:

http://www.spotlightonspend.org.uk/Lite/996/Royal+Borough+of+Windsor+and+Maidenhead

Public Sector News is supplied by ALGIS in LARIA (The Affiliation of Local Government Information Specialists in Local Authority Research and Intelligence Association) which represents information professionals providing information services to local authority staff and elected members. Jane Inman is currently Chair of ALGIS. ALGIS welcomes anyone with an interest in public sector information who feels they would benefit from being part of a community working in the same area. For more information go to http://www.algis.org.uk

Meeting Reports

Librarian's Information Literacy Annual Conference 2010

Conference Report by Sharon Lawler

This year's Librarians' Information Literacy Annual Conference (LILAC) was held in Limerick, Ireland – the first international location for what has become a popular date on the conference calendar.

As a student currently studying part time for an Information and Library Studies qualification at the University of Strathclyde, library conferences such as LILAC are usually too expensive to consider attending. However, thanks to UKeiG's generous support of my place, I was able to book a place and make my way to Ireland.

As well as studying, I work at the University of Glasgow, so I was particularly interested in sessions that addressed information literacy in Scotland, as well as those focusing on research and school libraries. As a copyeditor for the *Journal of Information Literacy*, the conference was also an opportunity to discover more about what was happening more broadly in information literacy to inform that role.

It was a packed schedule over three days, including a preconference session on the first morning, before the official opening. The conference was broken up into five themes: Making Connections; Developing the IL Practitioner; IL and Research; Innovative Practice and Measuring Impact.

Monday started with a busy preconference stream. As this had booked up quickly, I took the opportunity to start slowly and took the time to sit in on some quieter sessions and check my plan for the rest of the day. The Strand Hotel was already buzzing when we got there and the LILAC team, resplendent in their official T-shirts, were busily registering librarians from across the globe.

After the main conference was opened by the Minister for State for Lifelong Learning, Sean Haughey, the first keynote was courtesy of Tony Durcan, Head of Culture, Libraries and Lifelong Learning at Newcastle City Council. It was interesting to hear him describe the public library as "a pillar of interaction for citizens, and not just to read" and also that in Newcastle,

Internet access is not restricted to library card holders. In the discussion, it was noted that there was a lack of research focus on information literacy in public libraries, and that there were few if any public librarians attending the conference.

One session that did address information literacy in public libraries was John Crawford's session on employability training in public libraries. An interesting sign-off from the Scottish Information Literacy Project, which is currently being wound up in its present form, it highlighted that the fact that more work still needs to be done.

After the main thrust of the day, the evening heralded the networking evening at Limerick Institute of Technology. This was a great chance to meet up with other delegates, and we were treated to an eclectic mix of Irish dancing and a musical interlude from some members of the LILAC committee. Then it was back to the hotel to prepare for a packed day two…

A bright and early start on Tuesday with a 9am keynote from US researcher Professor Karen Fisher from the University of Washington Information School. Her research focuses on how some people (known as lay information mediaries or LIMs) seek information on the behalf of others. Fisher maintains that the "prevailing paradigm focuses on individual users" and that more research is needed to further understand social information literacy.

A full day of sessions followed on a wide range of subjects. The Pecha Kucha (Japanese for chit-chat, pronounced "pe-chak-cha") session featured various presenters, each given 20 slides and 20 seconds for each slide to present their topic. It was a fast-paced session and offered a taste of a wide range of subjects, particularly those that you may not have chosen for a full session, but without the opportunity for questions.

After a full day of sessions, it was time to head back to the hotel and get into my glad rags for the conference dinner, which was held at Dromoland Castle, a beautiful venue. After dinner, it was time for the awards ceremony and the following awards were presented:

- Information Literacy Practitioner of the Year: Anthony Beal, West Yorkshire College

- Honourable mentions: Andrew Walsh, University of Huddersfield and Nancy Graham, University of Birmingham

- Irish Information Literacy Award: Wexford County Library

Tuckered out, I headed out for the bus back to the hotel. Those with more energy stayed for the customary LILAC disco and much dancing ensued!

Sessions at 8:45am made Wednesday morning the earliest start of the conference. A busy morning included a session on information literacy support in school libraries, based on research from the School Libraries Group of CILIP, which is slated for publication in May this year, and an honest look at the training of users in the new British Library catalogue.

An interactive session followed, led by Lorie Khoda and Alison Brettle from the *Journal of Evidence-Based Library and Information Practice,* on interpreting information literacy research papers. This was really useful and we all came away armed with new resources.

Rounding off the conference was final keynote speaker, Dr Ralph Catts, Senior Research Fellow at the University of Stirling. Focusing on the academy, he spoke of the need for a large-scale research project to make a case for information literacy. In addition, he took the opportunity to summarise his thoughts on the conference and to critique the papers he had heard.

Catts admitted his aim was to be controversial and challenging, to encourage debate. His words were certainly thought-provoking. His idea of approaching educational researchers for assistance didn't necessarily take into account the research experience that some librarians or information departments already have, as some commentators pointed out, which he took on board. The lack of sabbatical time available to enable librarians to conduct research was also mentioned in the discussion. I thought this point may go some way to explain the issue of research in public libraries – if academic librarians struggle to get time off for research, then how will public and other sector librarians make their case?

All in all, LILAC was a great opportunity to meet librarians and learn more about information literacy, particularly in relation to higher education. As I work in higher education, this was highly relevant for me. However, as the keynotes by Durcan and Fisher in particular showed, information literacy is just as important in other sectors, including public libraries.

Social media also played a key part in the conference, particularly Twitter. There was an active twitter feed (#lilac2010) during the conference, to which I contributed (although not as much as I would have liked! Am now investigating smartphones….) and even a #lilacenvy thread! Twitter also played a part in connecting delegates before the conference, easing the pressure of going to a large conference without knowing anyone, for the many of us who were new to the conference.

Sharon Lawler is a postgraduate student, Information and Library Studies, University of Strathclyde

Current Awareness

Column editor: Clare Sinclair

Do you feel news from the world of information passes you by? That you can't keep up with the rapidly-changing e-publishing sector? Don't worry – we have solved the problem for you. This column is created by a team of information specialists based both in the US and the UK. They check a range of leading periodicals to make sure you are kept informed about the latest developments of relevance from them. Subjects covered include e-publishing, information access and retrieval, electronic publishing, preservation and virtual libraries.

E-publishing/legal issues

Cairns, Michael. **A Database of Riches: Measuring the Options for Google's Book Settlement** (22 April 2010)(http://www.scribd.com/full/30334705?access_key=key-23rh5w2lwcdmcmzph2k4) – Everybody has been waiting patiently to see if Judge Chin will accept the proposed amended settlement of the lawsuit brought by authors and publishers against Google, but with his recent confirmation to the Court of Appeals, it is unclear if Judge Chin will decide himself or turn the whole mess over to a new judge. In the interim, we have to continue to make educated guesses about what the Google database, if it is approved, will look like. In this new report, Michael Cairns, a former president of R.R. Bowker and a publishing consultant, who earlier tried to calculate the number of orphan works in the Google database, turns his attention to the business model for the institutional subscription. It is the first public analysis that tries to estimate how much a yearly subscription might cost. One can question his assumptions about likely subscription prices or whether the market penetration he anticipates would meet the settlement's requirement for wide acceptance of the product, but I was struck by his calculation of how relatively low subscription prices for the database of out-of-print books would generate huge windfall profits for the rights holders. He estimates $260 million in total revenues per year, 70% of which, or $182 million, would go to the rights owners. It suggests that a low subscription/high market penetration business model, rather than the high subscription/limited sales model normally adopted when marketing to libraries, would generate the greatest revenues. - PH*

Anderson, Ivy. **Model Language for Author Rights in Library Content Licenses** *Research Library Issues* (269)(April 2010)(http://publications.arl.org/pageview/rli/s68n7/12). – University authors are a bit like people who accept software click-through licenses: they blithely sign

their publishing agreements, assuming that they retain all the rights they want, and don't bother reading the fine detail to see if they have actually signed away their immortal souls. Librarians know better, and so have been at the forefront of initiatives that encourage authors to amend their publication contracts and adopt campus-wide open-access mandates that automatically preserve some important rights for authors. An ARL ad hoc working group is at work on a third, very promising approach. Since libraries already negotiate with publishers licenses for content (licenses that are worth hundreds of thousands of dollars), why not require as part of that content license that institutional authors retain needed rights, regardless of what their publication agreement might say? This article presents the group's draft model license that does just that. It is a clever way of exerting libraries' market power, and if widely adopted, could have an even greater transformative impact than the so-called "green open access" campaign. Comment on the draft now, and start thinking about how to include it in future content license negotiations. - PH*

Information retrieval

Brynko, Barbara. **The Power of the Semantic Web**, *Information Today* 27 (5) (May 2010): 10 – Overview of some work being carried out by the National Network of Scientists in the US on the creation of a semantic-Web application offering access to their research. Explains some of the principles behind semantic technologies including construction of ontologies.- MDS

Security

Shapiro, Stuart S. **Privacy By Design: Moving From Art to Practice** *Communications of the ACM* 53(6)(June 2010): 27-29. (http://cacm.acm.org/magazines/2010/6/92470-privacy-by-design-moving-from-art-to-practice/fulltext) – What with all the privacy problems Facebook has been having, a quick reading of this article might be in order. The argument here is that you start out with principles of security and privacy, building them into the system from the beginning and not as an afterthought once the product is nearly complete. Standards still need to be developed, particularly for privacy, but the author ends the piece arguing for "a firmer grasp of the obvious". - LRK*

Social networking

Lazar, Bart A. **Drafting social networking policies**, *Information Today* 27 (5) (2010): 20 – Useful follow-up to an earlier article on what organisations should think about when drafting policies for staff use of social networking sites. Although it is written from a US perspective where, it could be argued, the dangers of litigation are more than in the UK, there is some sensible advice such as the importance of disclosure by authors of blog posts and commentaries.- MDS

Virtual libraries

Hastings, Robin. **A Year With Google Apps** *Computers in Libraries* 30(4)(May 2010): 12-16. – At a time when many libraries are struggling to make ends meet, an article such as this that may save libraries money is welcome indeed. In this frank piece, Hastings takes us through the steps she took to transition her library staff from Microsoft Exchange and Office to Google Apps. In particular, it helps the rest of us a lot for her to note what they were able to migrate successfully (almost everything) and what they weren't (not much, in the end). Some things to note: going from using folders to manage email to using labels can be a conceptual and behavioral barrier, although certainly not an insurmountable one. Hastings wisely notes the importance of training in a transition like this, and another lesson was to not force people to use the technology as you would use it (they'll find their own way to use it). There are many libraries that should consider this path, and this article is a great place to start in looking into this option. - RT*

* Abstracts identified with an * next to the author initials, are drawn with permission from Current Cites, the monthly publication distributed electronically by a team of Librarians and Library Staff and edited by Roy Tennant: http://lists.webjunction.org/currentcites/

Contributors to Current Cites:

Charles W. Bailey, Jr., Keri Cascio, Alison Cody, Susan Gibbons, Peter Hirtle, Leo Robert Klein, Karen G. Schneider, Roy Tennant

Other Contributors:

Andrew Brown, Virginia Havergal, Martin de Saulles, Carrie Sherlock , Ann Smith, Matt Treadwell; Donald Tait; Nina Whitcombe

Books of Interest

Here are some book titles that might be of interest to UkeiG members. The list isn't exhaustive; members are welcome to suggest other books for review. If you are interested in reviewing any of these titles, or if you are publisher and you have titles that you think would be of interest to the UKeiG readership, please contact Michael Upshall (*michael@consultmu.co.uk*). A copy of the book is provided free for you to review.

New releases from Facet

Copyright and E-Learning: A guide for practitioners by Jane Secker

Fully supported with a broad range of practitioner case studies and further sources of information, this book contains a practically based overview of current and emerging copyright issues facing those working in the field of e-learning. The book is relevant for anyone working in the education sector from school level to higher education, and those developing learning resources in commercial organizations and the public sector including libraries, museums and archives, and government departments.

Building Your Portfolio, 2nd Edition: The CILIP guide by Margaret Watson

This 'how to' guide provides practical and clear advice on producing a successful portfolio. This revised edition takes account of the changes to the regulations for Certification; the new single route for Certification candidates, revised assessment criteria and a new personal statement template. Helpful new case studies reflect these changes.

Handbook of Art and Design Librarianship by Amanda Gluibizzi and Paul Glassman

Put together by an international team of contributors, this essential handbook examines methods of innovative librarianship in art and design libraries throughout the world.

The Tech Set Series Editor: Ellyssa Kroski

The Tech Set are ten practical fast-paced guides of current thinking and best practice for a wide range of library-related tech trends. Each title in the series is also accompanied by an author wiki and podcast.

UkeiG Training Sessions Update

The training programme for 2010 is now complete, with two courses set for October and November on making search work and legal issues and Web 2.0. These two have been put on by request from the membership. If you like would like to attend please book early to avoid disappointment as we know they are going to prove very popular.

Training courses for 2011 are now being planned. If you have a course you would like us to run or you would like to run for us, or if you have a venue you would like us to use or a suggested location do get in touch with the training coordinator, Tracy Kent on t.k.kent@bham.ac.uk or 0121 414 3918.

Our courses are becoming ever more popular as they provide value for money, are practical and meet the needs of our members. The training programmes, together with our factsheet programme, are a major member benefit and one we would encourage you all to make use of. More details on the revised training website at http://www.ukeig.org.uk/training.

UKeiG

UK eInformation Group

UKeiG AGM and Annual Seminar

16th June 2010

Getting the source out of the bottle: Practical Implementation of open-source Applications in the Information Sector

Open Source has many myths and notions about what it can and cannot do for an information organisation – whether it's saving you pots of money, or that it simply comprises products put together by hacker-wannabes. This Seminar seeks to explore what it means to implement OS applications by providing real, live, examples taken from across the information sector, so that you can assess for yourself where the advantages and disadvantages lie for your organisation.

Come along and share your experiences

http://www.ukeig.org.uk/blog/2010/03/ukeig-annual-seminar-getting-source-out.html

Press Releases

A time of transition: RIN annual review 2009

The RIN's second [annual review](#) has just been published and covers our project work and other activities during 2009, and gives you a preview of our work for 2010.

The RIN supports the development of effective information strategies and practices for the UK's research community. Through conducting research, providing guidance and promoting innovation, we aim to meet the needs of researchers, institutions, funders, information professionals and everyone who plays a role in the research information landscape.

The theme of transition ran through much of the work that we undertook during 2009.This is a time when everyone involved in research information is having to adapt and respond to changes in policy, practice and culture. Our major projects covered the use, value and impact of e-journals, gaps in access to information resources, case studies of information use in the life sciences and a study on how bibliographic records are generated distributed and used. We also undertook and produced a range of other projects, events and publications.

These are exciting times for those with a stake in the research information landscape. We need to be ready to embrace these opportunities and prepared to address challenges and issues that arise. Providing advice and guidance to the research community continues to be a major part of our work.

Referencing made easy

17 May 2010 – A new piece of software to help students and universities manage academic references more easily has been developed by the Open University and its partners.

Funded by JISC, and made in collaboration with RefWorks, the web-based research management, writing and collaboration tool, the OU has developed MyReferences to simplify referencing.

Lawrie Phipps, programme manager at JISC, said: "Writing up references can be a really time-consuming activity for the student and researcher – but it's essential to avoid plagiarism, develop robust scholarship and lead the reader to other relevant work. This resource is one aspect of a huge investment from JISC which aims to spark innovative solutions to issues like

referencing that affect people working up and down the country, in order to inspire and help other UK colleges and universities."

MyReferences integrates tools from the widely-used RefWorks reference management software into the Moodle virtual learning environment which is used in academic institutions across the UK to support online learning.

Nicky Whitsed, director of library services at the OU, said: "For students, this means helping them to adopt good practice in organising and managing references and using citations appropriately in their work. And for universities this new and easy reference management system will help to tackle plagiarism by making referencing easier; encourage information literacy; and help libraries to improve reference management for courses".

The MyReferences software is open source, making it free for any institution to download, tailor to their own needs and integrate into their own learning environments. The software is part of the 'technology enhanced learning supporting students to achieve academic rigour' (TELSTAR) project.

There are already a number of general referencing tools available to students, but MyReferences takes the usability of these tools a step further by integrating them into online courses so the materials students commonly need to reference are already available in the format they need. Students simply select the sources they need to reference, the referencing style their institution requires and then copy and paste the result into their assignment.

See a visual demonstration of MyReferences at
<http://www.open.ac.uk/blogs/telstar/telstar-quick-demonstration/>

Find out more about events planned to share information about TELSTAR and developments in the field of reference management at <http://www.open.ac.uk/telstar/event>

Digital Libraries a la Carte

From 26 to 30 July 2010, Ticer's international summer course "Digital Libraries à la Carte" will be held at Tilburg University, the Netherlands. The course is officially recommended by JISC.

The programme of this annual event for librarians and publishers is completely renewed. The course modules cover:

- strategic developments and library management

- the library in the scholar's workflow and research data

- libraries - partners in teaching and learning

- mobile technologies in education and library

- Web 2.0 and linked data in libraries

The course website can be found at www.tilburguniversity.nl/ticer/2010. The course brochure can be found at http://www.tilburguniversity.nl/ticer/2010/Ticerbrochure2010.pdf Those registering before 1 May 2010 will get a 150 euro discount.

Oxford Bibliographies Online

www.oxfordbibliographiesonline.com

In recent years, the explosion of original research and its increasing accessibility through new technologies is a blessing for scholars—but also a tremendous challenge. For faculty, students and librarians querying new areas of research, the most pressing challenge for researchers at all levels is sifting through and managing the immense amount of material available online. "With the growing ease of digital access and dissemination, scholars and students are increasingly overwhelmed by the enormous volume of academic material available to them," noted **Nigel Portwood**, OUP's Chief Executive. "The need for responsible and sophisticated filters has become critical. *Oxford Bibliographies Online* represents a significant step forward in providing meaningful navigation through the vast body of research material accessible on the web."

Oxford Bibliographies Online—a series of intuitive and easy-to-use "ultimate reading lists" is designed to help users navigate vast seas of information. OUP's first online-only product goes beyond search algorithms to present bibliographies in several disciplines selected by leading subject experts and vetted by the highest professional standards.

The result of extensive interviews with librarians, numerous focus groups and market research, *Oxford Bibliographies Online* extend a centuries-long mission of supporting excellence in research, scholarship and education. *Oxford Bibliographies Online* will help scholars increase productivity, save time and elevate the quality of their work, while enabling them to more efficiently engage with platforms like Google Books.

- Classics, Islamic Studies, Social Work, and Criminology will launch *Oxford Bibliographies Online* (*OBO*), with an additional 10-12 subject areas (modules) scheduled to launch within the next year

- 50-100 entries per subject at launch (equivalent of a 4-volume print encyclopedia)

- Modules forthcoming in May 2010: Atlantic History, Renaissance & Reformation, and Philosophy

- Every subject area has an Editor-in-Chief and Editorial Board

- Each entry receives multiple peer reviews and board vetting

- Each Individual entry provides expert recommendations through key literature on a specialized sub-topic. It features an annotated bibliography broken into subjections, each preceded by context-setting text introducing the entries that follow

- All entries are selective, rather than exhaustive

- Entries provide cataloging of non-print material, especially online resources, critical to the entry

- Updated and revised quarterly, with 50-75 entries added per year to each subject area.

- Linking on multiple levels facilitates movement to other resources, whether online or available through a library's catalogue.

About *Elucidate*

Elucidate is the journal of the UK Electronic Information Group. It is published five times each year, in February, March, May, September, and November. It aim is to keep members up to date with developments in the digital information environment, as they affect professionals. The magazine is provided free to UKeiG members.

Notes for contributors

Elucidate welcomes articles or ideas for articles in the areas covered by the magazine. We are always on the lookout for feature writers, reviewers both for books and for meetings, as well as respondents to articles. Sadly, we don't pay contributors, but contributors retain copyright of their articles and can republish their articles elsewhere.

Brief for feature articles

If you are writing for *Elucidate*, please follow these simple guidelines:

About the members

Our membership comprises information professionals involved in the dissemination and/or delivery of digital content and services. Our membership base is two-thirds academic, one-third commercial, as well as some public libraries. A key benefit of the group is that meetings and forums provide "crossover" insight from one area to another: members see it as a way of keeping up to date in areas outside their core. Few other organisations provide this kind of cross-sectoral awareness. The focus of the group is the UK, in the sense that authentication concerns tend to be around JISC tools such as Athens and Shibboleth. But the issues of digital provision are of course global. The most popular training courses we run are on search tools – Google and others; e-books and how to deal with them. Other popular strands include intranets, content management, bibliographic software, and e-books.

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. Each article should be prefaced by a short summary (around 50 words) that can be used when displaying on public search engines an outline of the article, and to display on the non-member section of the website.

What to write

A key aspect of UKeiG is that it provides insight from one area to another – members see it as a way of keeping up to date in areas outside their core expertise. Because the membership is quite disparate, ranging from pharmaceutical information professionals to public librarians, you should not assume the readers are as familiar as you in the subject area you are writing about.

The most valuable viewpoint you can give is that of an end user. UKeiG is not a place for theoretical debate, but a forum where peers can share their experiences and understanding. So, if it worked for you, tell others. If it didn't, tell others why not.

How to submit

Please e-mail your article to michael@consultmu.co.uk. Articles should be delivered in Word or in an ASCII format. Images are welcome – 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 in their content and remain the copyright owner.

About you

Please provide a 10-20-word biographical summary about yourself to appear at the end of the article.

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.

Brief for book reviews

Book reviews are typically 500-750 words. Because UKeiG is independent of any book sales operation, it is not obliged to have favourable reviews. If you think a book is poor, then by all means explain why. Members and non-members alike are welcome to suggest books for review or to submit reviews.