Catalogue & Index

Periodical of the Chartered Institute of Library and Information Professionals (CILIP) Cataloguing & Indexing Group

Editorial

Welcome to Issue 160 of Catalogue & Index.

The main theme behind this issue was the idea of sharing everyone's knowledge and experience from any conferences, workshop or seminars attended in the last few months.

Helen Williams of London School of Economics attended the recent Talis open day covering all things semantic and how to apply these tools and structures in a library setting. Helen has included links to all the presentations from the day that covered a wide range of topics from making library resources web3 ready and the frameworks and schemas used in the Linked Data movement.

We thought it was about time to let you in on the behind the scenes of organising a CIG visit and Deborah Lee has furnished C&I with a great eye opener as to what is involved. Anyone interested in organising a visit for CIG members, contact the committee, Deborah makes it sound all so easy!

Antonio Jimenez describes the realisation that cataloguing is a dark art and the enjoyment one can receive from delving into a special collection and finding out more than you ever thought about one man and his interests.

I attended a fantastic free! two day workshop on maximising your online assets, in my hometown of Glasgow. It has given me new verve and hope for metadata practitioners and the need for understanding structure and rules is probably more prevalent than it has ever been.

This issue rounds up with a call to arms for anyone interested in getting involved with the Cataloguing & Indexing Group, without being a member of the committee, you know you really want to!

Penny Robertson, Editor

Contents

Talis Open Day: Linked Data and Libraries

Helen Williams , London School of Economics

'Linked Data and Libraries' was organised by Talis at the British Library on 21st July 2010 with the aim of introducing participants to the principles, practice and potential of Linked Data.

After a welcome and introductions by Richard Wallis of Talis, Zach Beauvais, editor of Talis' *Nodalities* magazine, gave a brief introduction to Talis and the world of Linked Data. Talis is involved in building next generation web applications and the purpose of *Nodalities* is to 'bridge the divide between those building the Semantic Web and those interesting in applying it to their business requirements'.
For further details see http://www.talis.com/nodalities/ Early issues of *Nodalities* addressed the possibilities offered by the Semantic Web while more recent issues contain articles which detail work in progress, reflecting the way in which the web of data has been emerging over the last two or three years.

The second speaker, Romain Wenz from the Bibliothèque Nationale de France (BNF), talked about their Linked Data work; the data bnf.fr project. This project is seeking to make library resources easier to find on the web by creating web pages which group resources round the concepts of works and authors. Using HTML to build the web pages the data will be gathered from both internal and external sources to create an information hub which can be browsed both by users and search engines. The library currently has eight different catalogues which will need to be internally interoperable for this to work so the project will require significant changes to library policies in order for it to be a success. As well as addressing issues of efficiency and interoperability the project needs to be viable in the Semantic Web environment. This means that data needs to be accessible in RDF (Resource Description Framework), and that URIs (Uniform Resource Identifiers) need to be maintained and to remain stable. The project is currently at the early stages, but will ultimately make retrieved data more valuable for users.

Once we had seen how the BNF are hoping to harness the possibilities of Linked Data, Rob Styles from Talis took us back to basics on Linked Data, RDF and SPARQL, with a clear presentation on these technical topics. He explained Linked Data as the technical practices employed to create a web of data. Today the web is a global network of linked HTML documents. As a place where anyone can say anything about anything it is a decentralised and unregulated environment. The documents making up this vast collection of 'human readable' knowledge and opinion are linked, but the links are not qualified and so there is no meaning in them.

The web of data is 'socially' still the same decentralised, unregulated environment, but in contrast to the web we are familiar with, it is a global network of linked statements. The links between these statements are qualified, which means that the data is machine readable instead of only 'human readable'.

The technology underpinning the web of data is RDF (Resource Description Framework), which Rob described as a simple and powerful data model specifically designed to work with the architecture of the web and define

relationships between things. It means that 'human readable' web page statements can be translated into machine readable statements and therefore can be analysed, searched and indexed in information retrieval. In RDF the resources involved, such as web pages, people or places, are named using URIs. Two people using the same URI are assumed to be describing the same thing and so when these are merged, they become one URI containing the properties of both datasets. Using data from many different websites this can create a vast amount of Linked Data. Rob pointed out that one does, of course, need to trust the sources from which the data is obtained, or to have some kind of internal validation process.

The final part of Rob's talk dealt with some more technical matters, covering the ways in which RDF can be written down using Turtle, N-triples, RDF-XML, RDF-JSON and RDFa, before moving onto SPARQL (SPARQL Protocol and RDF Query Language). This is a graph-pattern-based query language for RDF which allows searches to be carried out using the queries 'Ask, Select, Describe and Construct'.

In the final session before lunch we turned from theory to practice as Richard Wallis showed us some sites already using Linked Data together with the RDF data behind them. To pick up on just a couple of these examples, we looked at the catalogue of the Royal Scottish Academy of Music and Drama http://prism.talis.com/rsamd/ where we saw a standard catalogue interface with results and facets. What users do not see is that this is running on top of a Linked Data dataset which means that results pages can link out to other related items. In this case the links are to other items in the same data set, but a site such as BBC wildlife http://www.bbc.co.uk/nature/animals/ has taken this a step further. Building their web pages on Linked Data principles and using URIs as identifiers, they are able to pull in data from other well known and reliable sources rather than gathering and holding all the information themselves.

After lunch we heard 'Lightening Talks' from three of the Open Day attendees. First was Neil Wilson, Head of Bibliographic Development at the British Library, talking about plans for using BL data in new ways. He referred to the Government's 2009 commitment to the principle of opening up public data and facilitating greater use of Linked Data. Neil pointed out that an open data strategy has certain financial and legal implications such as loss of revenue from traditional bibliographic services and increasingly complex copyright issues, and also reputational implications in terms of whether the library is perceived to be 'keeping up'.

The BL is looking to go beyond traditional MARC data and has already developed the capability to support non-library metadata standards as used by the wider web community. In May 2010 they hosted a Linked Data workshop to address technical challenges, standards and prototypes, and they are now doing some further work with JISC and participating in an Open Bibliography project with Cambridge University Library.

Next, Sally Chambers from the European Library spoke about the Europeana Data Model. The European Library already offers access to the resources of 49 libraries in Europe and Europeana's aim is to make digital information from libraries, museums, archives and audio-visual collections within Europe accessible. Using the

European Data Model, which has been produced to facilitate participation in the Semantic Web, Europeana is working on a framework for connecting, collecting and enriching metadata to enable the browsing of these digital objects in new ways.

Finally Felix Ostrowski from the Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen spoke about building a Linked Data based index of library institutions. Currently the data held in the various sources which might inform such an index is only very basic or is outdated. Up to date or detailed information is available on library websites, but these are usually only 'human readable'. There are still decisions to be made surrounding copyright issues and modelling decisions, but if libraries can be encouraged to use RDFa in their web pages, then their data is turned into a machine readable interface and the quality of the data which could be harvested to a centralised source is greatly enhanced.

Rob Styles followed the 'Lightening Talks' with a presentation showing us how MARC data can be transformed into RDF. Each RDF property (for example, author, title, subject) is labeled and given an http URI so that it can be an RDF link on the Web. Authority and bibliographic data, which are kept separately in MARC, are meshed together in RDF through sharing URIs, thereby bringing data together. The aim is to create, in the web, models of data which identify all the things the MARC record describes but using different technology. For librarians some of the important developments are Library of Congress Subject Headings as Linked Data http://id.loc.gov/, the Virtual International Authority File (VIAF) http://viaf.org/, and the work in progress to publish Resource Description and Access (RDA) vocabulary as RDF http://metadataregistry.org/.

Antoine Isaac spoke about the W3C Library Linked Data Incubator Group. This is a group which is running for one year, not to make recommendations about Linked Data, but rather to look at innovative ideas in this area. Antoine highlighted that though libraries have a long history in producing metadata, usually of a high quality, this metadata often remains locked in bibliographic records. This group hopes to help increase the global interoperability of library data on the web by bringing together people who are involved in the Semantic Web, building on existing initiatives and identifying areas of collaboration for the future. As well as gathering case studies of success stories the group will identify relevant data models, vocabularies and ontologies and consider the need for new standards, guidelines and best practices.

Before an open discussion to end the day Richard Wallis gave an overview of the Talis platform, showcasing the cloud-based data storage on offer from Talis as a Linked Data publishing platform.

The day was very informative and it was exciting to hear about all the Linked Data projects that are already in progress and to consider the extent of future possibilities for libraries.

Videos from the day can be seen at:

http://blogs.talis.com/nodalities/2010/08/linked-data-and-libraries-almost-like-being-there.php

This article is based on the presentations given at the Talis Open Day at the British Library on 21st July 2010 and credit should be attributed to Richard Wallis, Zach Beauvais, Romain Wenz, Rob Styles, Neil Wilson, Sally Chambers, Felix Ostrowski and Antoine Isaac.

**Post notes of interest & feedback at the CIG blog or
email any member of the
committee:
http://communities.cilip.org.uk/blogs/catalogueandindex/default.aspx**

CIG Conference 2010, presentations available at:

http://www.ukoln.ac.uk/events/CIG/2010/conf-exeter/programme/

Issue 161 of C&I will host articles created by speakers at CIG Conference 2010

Email the Ed, any idea for themes or topics you want covered in next year's journal, get in touch!
e: penny.robertson@sqa.org.uk

From cricket to art history: attending and organising a CIG visit

Deborah Lee, Courtauld Institute of Art

As a cricket agnostic, I wasn't the obvious candidate for a trip to the MCC library. However, as a fully paid-up cataloguing geek, the opportunity to spend an afternoon looking at a home-grown classification scheme as well as a guided tour of a historic building was too much to resist. As the keeper of our own classification scheme, I was excited about the chance to see how another library manages their scheme. From any perspective, the library was a fascinating place – there is a TV above the enquiry desk and comfy chairs! – but to a classification obsessive, it was highly educational. The issues involved in classifying materials about cricket venues, chimed heartily with my own experiences of classifying by concert venue and by art gallery.

The visit finished with a tour of some of the more famous buildings which make up the MCC, including the England team's dressing room, the Long Room and the members' bar. My lack of cricketing expertise did not diminish the pleasure of tea and nice biscuits while chatting to cataloguers in beautiful surroundings. I may not know my Flintoff from my Pietersen, but standing in the England dressing room trying to imagine what it would feel like to be waiting for a call down to the famous pitch was a very moving experience.   So while eating those nice biscuits, surrounded by friendly cataloguers, I ticked that box on the feedback form saying that we would be interested in hosting a CIG visit.

So, a few weeks later ...

Around Easter, I received an email asking whether the Courtauld would be willing to host a CIG visit. The email included various helpful details about the process, and making the initial arrangements was very straightforward. First, I outlined to the CIG contact what areas our visit could cover.   Next, I discussed possible dates for the visit with our head librarian and asked our CIG contact to select their preferred date. A small hitch occurred when we realised that one of our key cataloguing team would be on leave on our selected day; however, our CIG contact checked swiftly with her fellow CIG colleagues that our new proposed date would be acceptable.

With the all-important date now fixed, the Courtauld cataloguing team discussed the contents of the visit.   We decided to focus on three facets of our work that were specific to our library: cataloguing art-historical formats, the cataloguing workflow in a small university library and our historic classification scheme. My colleagues then volunteered – with only a little coercion! – to talk about one of these areas, and we all discussed what form each talk might take. As well as the cataloguing talks, the visit would start with a tour and history of the library, and finish with an opportunity for tea, biscuits and networking.

The next stage of organisation involved writing the official blurb for the visit and publicity. As CIG provide a helpful template, writing the aims and objectives for the visit was very straightforward. The CIG contact then organised the publicity for the visit, and as I had selected to be the contact for attendees, I then waited for the replies to trickle in. I found it fascinating to see the variety of attendees' geographic locations and library-types. A week before the visit, I wrote a timetable for the day, to ensure that my colleagues knew where they had to be and when. In addition, my colleagues and I put the finishing touches to our small exhibition.

For my colleagues and myself, the visit provided an excellent opportunity to discuss an aspect of our work with colleagues from other institutions. Preparing for such a talk enabled all of us to ponder the mechanics and processes of what we do, and what we do that may be different from other libraries. As the person giving the classification talk, I found it an invaluable opportunity to concentrate on specific classification issues in our library and to find visual examples of these issues. For some colleagues, their talk was their first experience of giving a formal presentation to cataloguers outside of the library; they found it a very worthwhile and useful experience. From the organising perspective, I relished the opportunity to arrange a visit which focused on cataloguing and classification. The support given by CIG was faultless, particularly that provided by our CIG contact, Katrina Clifford. I can thoroughly recommend organising a CIG visit: it provides an opportunity to gain insights into your own cataloguing and classification activities while sharing thoughts and ideas with colleagues from across the United Kingdom.

The Pamphlet Collection of Major J. Fairfax-Blakeborough

Antonio Jimenez, York Minster Library and Daryl Green, former Library Assistant

In 1976, York Minster Library came into possession of a 20-volume set of bound pamphlets detailing Yorkshire life in the late 19th and early 20th century. The collection was bought from the estate of John Freeman Fairfax-Blakeborough (J.F-B) in the year of his death. These volumes of pamphlets were numbered and indexed and then left in our Upper Stacks for storage.

32 years later, in the winter of 2008, the collection began to resurface. Over time, the collection had been broken and, at first, we could only find six volumes, averaging about 25 pamphlets each. We began cataloguing not knowing that another 12 volumes awaited us in the depths of the Upper Stacks. As we would finish cataloguing a volume we would have a look upstairs just to make sure that no others could be found, and undoubtedly one of us would return with another few volumes, until at last we have found all 20 of the original purchase.

In this long process of cataloguing nearly 400 pamphlets, we became quite familiar with the character of J.F-B. This collection is not just groupings of pamphlets bound together, they contain J.F-B's handwritten comments, newspaper clippings inserted in relative pamphlets, personal correspondences that have been left by various authors and Yorkshire folk and cards and miscellanies from other racing fans and Yorkshire dialect specialists. This has painted a very intimate portrait of J.F-B for us cataloguers, a portrait which his memoirs, published by his son Noel, does not offer.

Before opening the first volume, we did not know any details about the life of J.F-B. We only got to know him through our cataloguing work, discovering more and more about his experiences and tastes through his personal collection.

While filling the cold MARC fields of our Aleph-based online catalogue, we found out that he served in World War I and contributed to *The Mudhook, the journal of the 63rd (R.N.) Division*, where J.F-B used his skills as a horse-racing reporter from Yorkshire to keep his division up-to-date and even offer a bit of humour. We found that he was passionate about and had an important reputation within the horse racing community, including books like *Horse facts and fancies* and *Post and paddock pars and poems*. We found J.F-B's passion for Yorkshire life and dialect, kindled by his father's stories such as *Mrs. Waddleton advertises for a "sarvant lass" (Being an old Yorkshirewoman's views and experiences on the servant question)* by Richard Blakeborough, and other series in which J.F-B would edit and release posthumously and later stories such as *Lizzie Leckonby* which he would write himself. Of course, he enjoyed the pleasure of the good northern life, collecting titles such as *The story of the Wensleydale cheese* by T.C. Calvert and guidebooks to such small places as Osmotherley, Bedale and Ampleforth.

The cataloguing of this eccentric collection finished in August 2008 and the pamphlet collection of J.F-B is now reunited on the shelf for the first time in 30 years, available to the public in our reference collection. References to this entire collection are available on our online catalogue and provide an invaluable resource to our readers interested in Yorkshire history.

Above all, it showed us that cataloguing is far more than a mechanic skill. It provides a privileged insight into unique material and requires research and an interest that goes beyond the strict observation of AACR2 rules. Behind J. F-B's pamphlet collection there is a life full of experiences, feelings and knowledge that we would not have discovered without our cataloguing job. Now the information is structured and ready to be found by every one. When reading the pamphlets we hope our readers will feel part of the excitement we felt while typing on that blank ALEPH screen.

Maximising online effectiveness of your online presence workshop

Penny Robertson, Scottish Qualifications Authority

In this age of constant cutbacks, efficiency savings and jobs disappearing it was with great excitement to receive an email from the Strategic Content Alliance and JISC NetSkills offering a **free** two day workshop in my hometown of Glasgow. I duly forwarded to a colleague after I'd read the full content for the days and not to mention the location, a hotel just on George Sq., renowned for it's catering prowess!

The workshop was billed as intending to address the importance of optimising your online resources for maximum benefit and it really delivered. The first part of the workshop looked at how to improve search and retrieval of online resources, covering basic common sense principles in a very practical and logical way. The attendees were asked to create a mind map of all the key points that they would take into consideration for their own web resources. Each table did a small presentation that I thought was a great ice breaker for the workshop. C&I readers will be glad to know that every table mentioned metadata, structure and consistency of approach!

I was lucky to be at a table with a real mixture of both experience and sectors. I found this very beneficial to hear how others deal with their online resources and how to align these with business objectives and deal with constraints of IT departments.

My organisation, Scottish Qualifications Authority (SQA) is very content centric and I wanted to find out about how to develop a content strategy, how to apply this in business processes and what should be our main points in a social media and content strategy. Some astounding stats were given from brandinfiltration , if Facebook was a country it would be the third most populated in the world, ahead of the US or 95% of companies use LinkedIn to find employees. The presenters, George, Hanna and Christine really gave us great food for thought but really, at the end of the day, metadata truly does make the world go round.

My main interest in attending was finding out more about RDF and how to apply this in our organisation, the day really gave us a practical insight in to where to start. George gave, in my opinion, probably the best description of RDF triples I've ever heard. He explained how RDFa (RDF in attributes) can allow RDF to be included in html files using formally defined attributes in the <span> block and how vocabularies used are referenced in the <head> tag, really useful information for a newbie! Myself and my colleague are already thinking of a small pilot project looking at structuring question papers in RDF.

The workshop was excellent and well worth 2 days away from the daily grind. It was inspiring to hear that metadata is as necessary now as it has ever been and probably more so with the increasing content available on the web, it's just getting your head round some new languages and processes.

To finish, I'd like to include a link to Gary Flakes demo of pivot at TED conference, showing how structured data can become the bedrock of some truly awesome data visualisations.

My colleague was interviewed by the team at the end of the event and he's kindly given me permission to link to his five minutes of fame , a good sum up of the event.

To find out more about the workshops and check out the presentations, visit: http://scamore.eforum.org/cgi-bin/default?section=information

Six ways to get involved (without being on the Committee!)

Committees not your thing?

The good news is that there are plenty of ways of becoming involved with CIG that don't require that you sit on the committee.

1. Come along to an event or a visit, learn something new and network with colleagues working in jobs like yours.

2. If you have not seen an event that takes your fancy - how about suggesting one? Even better, help to organise it with the help and support of the Committee.

3.   Have you been to an interesting event, or discovered a fascinating special collection? Do you think that your library or information unit does something, however small, particularly well? If so, then let others know about it either via this blog (for details of how to post on the blog, see our Blog Policy) or through our journal, *Catalogue & Index.*

4. Write an article for *Catalogue & Index* – look out for calls for papers from the Editor, Penny Robertson, for specific themed issues or contact her if you have an idea for an article or an issue theme. Penny can be reached at: Penny.Robertson@sqa.org.uk

5.   Review a book for *Catalogue & Index* – keep an eye on the blog for calls for reviewers from our Book Reviews Editor, Neil Nicholson.

6. Contribute to the debate on professional issues by raising a topic on the blog - help to generate a cataloguer's perspective on the issues of the day by putting in your two pennies worth!

The Group is its members and the more of you that get involved, no matter in how small a capacity, the more lively and varied our activities can become.

So if you feel like dipping a toe in the water but don't know where to start, contact anyone on the Committee.

We are waiting to hear from you!

http://communities.cilip.org.uk/blogs/catalogueandindex/default.aspx

Next-Gen Library Catalogs.
Marshall Breeding. London: Facet Publishing, 2010. (*The Tech set* ; 1).
x, 138 pp. ISBN 978-1-85604-721-0.

The author of this book is well known as the foremost authority on this topic, so it makes sense to have him start off this impressive, ten-volume series of books which deal overall with technology and technological influences on librarianship. This particular topic has been of intense interest in the profession for a number of years, as libraries have tried to grapple with clunky, proprietary integrated library systems (ILSs) that the users have abandoned as a discovery tool since the mid-1990s.

First of all, each book in this series has three major components: the book itself, a companion wiki found at http://techset.wetpaint.com, and author podcasts. There is also a Facebook (http://facebook.com/nealschumanpub) and Twitter update pages (http://twitter.com/nealschumanpub). These books are meant to be a set of practical guidebooks specifically for libraries and librarians that are looking to implement and use cutting-edge technologies.

After a Foreword by the series editor, the author has a Preface that succinctly describes the organization and audience for the book. Chapter 1 is the Introduction, which provides a brief history of the evolution of the library catalog, why the current status quo of legacy (i.e. proprietary) online public access catalogs (OPACs) have never worked well for users, and how a new genre of software labeled "next-generation library catalogs" provide access to a vast range of library collections that are not currently accessible in the legacy OPACs. Chapter 2 goes through the process of planning for a next-generation catalog: crafting a user interface, understanding product options, selecting and procuring a discovery interface, and user-centered design. Chapter 3 then deals with implementation issues: integration with a current ILS, extracting and importing data, cataloging and metadata issues, integrating article-level and enriched content, monitoring and maintaining the system, and keeping track of emerging trends. Chapter 4 gets into the marketing of the new interface: using public relations experts, web positioning, interface branding, how to market to library personnel, and employing search engine optimization. Chapter 5 examines best practices through discovery interface effectiveness techniques, and how to assemble a toolkit. Finally, Chapter 6 discusses assessment techniques such as benchmark patterns and trends, measurement and analysis tools, and measuring the impact of discovery. There is an annotated bibliography and index at the end of the book.

The author has written a number of articles that seem to culminate in this practical handbook of information on the topic of next-generation library catalogs. It is written in clear, understandable language, and would be a valuable asset for any library or librarian examining the range of possibilities for planning and implementation of such software.

Bradford Lee Eden, Ph.D.

Associate University Librarian for Technical Services and Scholarly Communication

University of California, Santa Barbara

Santa Barbara, California

eden@library.ucsb.edu

Catalogue & Index is electronically published by the Cataloguing and Indexing Group of the Chartered Institute of Library and Information Professionals (CILIP) (Charity No. 313014)

Subscription rates: free to members of CIG; GBP 15.00 for non-members & institutions

Advertising rates: GBP 70.00 full-page; GBP 40.00 half-page. Prices quoted without VAT.

Submissions: In the first instance, please contact the Editor: Penny Robertson, Information Architecture Manager, Scottish Qualifications Authority (SQA) e: penny.robertson@sqa.org.uk

For book reviews, please contact the Book Reviews Editor: Neil T. Nicholson, Cataloguing & Metadata Services team Leader, National Library of Scotland, e: n.nicholson@nls.uk

ISSN 0008-7629

CIG website: http://www.cilip.org.uk/specialinterestgroups/bysubject/cataloguingindexing

CIG blog: http://communities.cilip.org.uk/blogs/catalogueandindex/default.aspx

Tags from the CIG blog on specific areas of interest:

authority control book reviews Catalogue and Index cataloguing CIG activities CIGS classification committees and working groups conferences Dewey digitised material Dublin Core events folksonomies linkblog MARC metadata news RDA Semantic Web social software standards taxonomies UDC