

Catalogue & Index

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

Issue 158

Editorial

Welcome to **Issue 158** of Catalogue & Index. This issue covers the main topics of authority control and indexing. With the onslaught of digital information, the need for a controlled semantic value of citing people and places ever increases. This issue demonstrates how our profession is, and can be, at the forefront.

Information professionals are the practitioners of a tradition that spans hundreds of years but is ever more prescient in this time and age.

Helen Williams of the London School of Economics and Political Science gives us an insightful and practical guide in how to plan and process any retrospective authority control project and how a balanced blend of human and technological intervention can reap great reward.

With the relatively recent change in MARC21 by MARBI to make the 440

field (that could be used for added entries for series) redundant in favour of using the 490 and 8XX fields, Colin Duncan, Inverclyde Libraries discusses the implication and includes some information on other library services and how they will deal with the change.

Kathleen Menzies, Researcher at Centre for Digital Library Research at Strathclyde University offers a precise

account of the many ways the online service Wikipedia uses the practices of classification, categorisation and meta-data to organise its site and allow varied access and navigation to its legions of users.

Hugh Taylor discusses the importance of the globally renowned NACO authority files, maintained by the Library of Congress. The file exceeds 7 million

records and 25% of new additions come from outside the US.

Keeping with the international flavour, Jennine Knight of the University of West Indies highlights both the strengths and weaknesses of pre and post coordinate indexing.



Alan Danskin of the British Library, Chair of CIG, takes us through some new and exciting developments in the world of authority control and the possible benefits and challenges facing the information sector. Elly O'Brien rounds off this issue's articles with an account of using OCLC's online WebDewey service.

Penny Robertson, Editor

Contents

2 - 3 Retrospective authority control, **H K Williams**

4 - 5 Series authority control and MARC21, **Colin Duncan**

6 - 9 Wikipedia and cataloguing, **Kathleen Menzies**

10 - 11 Cooperative name authority data – the LC/NACO Authority File, **Hugh Taylor**

12 - 16 Pre and post coordinate indexing: strengths and weaknesses, **Jennine Knight**

17 - 20 Spelling it all out: FRAD, ISNI, RDA, VIAF: automation and the future of authority control, **Alan Danskin**

21 - 23 Using WebDewey, **Elly O'Brien**

plus...

CIG authority control seminar report

Book reviews by CIG(S) members

We began considering an authority control project at the LSE Library at the beginning of 2006. By this time the loading of all retro-converted records was complete and an authority control group was convened in order to assess user needs. Authority control procedures had varied somewhat over the years, primarily because each library management system had offered different methods of verifying headings. As well as the legacy of system migrations, records had been imported from a variety of sources, and practices had become particularly unclear following the migration from Unicorn to Voyager in 2004. Only the Library of Congress Subject Headings file was purchased and so all new name authorities had to be authorised by manually importing the record from the Library of Congress. It became apparent that some staff were checking against existing entries to achieve consistency, others were importing Library of Congress records, while others were creating in-house authority records where none was available. This meant there were a number of variant headings in the catalogue.

The decision was taken to outsource a retrospective clean-up of all authority headings (subjects, series, and names) from a company who could also provide a regular ongoing check of the catalogue. At the same time, clear authority control guidelines were established for staff to ensure as few inaccurate headings in the catalogue as possible.

Tenders were assessed and the project was awarded to Marcive. Marcive would receive an electronic

copy of our catalogue to verify all name, subject and series headings against Library of Congress authority files using automated processes. Bibliographic records would be amended to contain corrected headings. We would receive corrected records and new authority files for loading as well as reports of unmatched headings. This makes the process sound simple, but we discovered various complexities along the way.

Preparing to send the file to Marcive required some in-house planning. In particular we needed to be aware that any changes we made to our bibliographic records while the file was with Marcive would be overwritten when their corrected data was supplied. We therefore excluded order records from the data we sent so that we could continue to accession books during the project. These records were sent to Marcive for checking after the initial data clean, as part of our ongoing services. We kept a spreadsheet of existing bib records requiring changes during the course of the project so that they could be corrected afterwards. This built up into quite a considerable amount of work because the project took longer than we had anticipated.

In May 2007 we exported approximately one million bibliographic records to Marcive and two weeks later received a test file of 10,000 records for checking. We checked one in ten of these, a very time consuming process. We undertook such thorough checking aware that we had exported our entire catalogue to Marcive, and that the methods which had been used on this sample would be used

on all our bibliographic records. Had we received records with errors or corrupted data back into our catalogue it would have a hugely detrimental impact on our users.

Our checking did reveal a number of queries which we submitted to Marcive. They provided a speedy and detailed response, though this indicated that we had higher expectations of the automated process than was actually achievable. Fortunately the things we had hoped would be corrected through automated processing appeared in accompanying error reports thus reassuring us that we would still be able to clean the catalogue to the degree we had originally intended, albeit that it would involve more staff time than we anticipated.

A few weeks later we received all our corrected data and accompanying authority records. Our IT department began loading files into our test server. Unfortunately the test server suffered under the strain of so much data and we had to wait for Ex Libris to carry out a regeneration of the indexes before we could proceed. After this it was unsurprising to find that loading one million bibliographic records and 500,000 authority records into the live server was not without problems either. Having started the process, IT estimated that loading the files and regenerating the keyword indexes would take 30 days because it was such a slow process. One option was to take the live server offline, but downtime is inconvenient to staff and students alike so this was not a particularly practical option. The

other option was to re-index in large batches at the end of the file loading. This meant that for about a week there would be inconsistencies in OPAC searches whereby the search facility used the old indexes but records contained new data. This seemed the most practical way forward, however, and 143 hours later all the data was loaded and the re-indexing completed just before the end of 2007.

As 2008 began it was time to think about the ongoing processes Marcive would be providing for the Library. We send files of new records to Marcive on a monthly basis and they clean them and send them back along with any necessary authority records. Supplied with this is a report of anything unrecognised or with multiple matches and therefore requiring human intervention. This is worked through by a member of staff to tidy up outstanding headings.

Running in parallel with this is a notification service whereby Marcive have a copy of all our authority headings and notify us if a relevant record is added to Library of Congress, or if there are changes to any of our existing authority records. These are dealt with through the Global Headings Change facility in Voyager. A member of staff can approve or disapprove changes and this is applied to all related records. This list is generated automatically by Marcive and we have found it requires manual intervention rather than automatically approving all changes, as some incorrect headings are suggested and we would not want these to be applied to all related bibliographic records. The Global Headings Change facility in Voyager was not as

straightforward to use as we envisaged. We discovered eventually that this was due to a bug in the system meaning that we were unable to link new authority records to related bibliographic records. We had to wait for Ex Libris to resolve this and in the meantime had to stockpile reports received from Marcive. Once this was resolved another bug meant we had 'orphaned headings' which would never clear from our list. After more work from Ex Libris this was sorted out and we were able to work on the backlog of reports.

In addition to these ongoing services from Marcive we still continue with our existing authority guidelines for in-house work. Authority work is far simpler with the item in hand as it prevents further work in terms of unmatched or possible duplicate headings which would often require retrieving the item from the shelf in order to correct the record at a later stage.

As well as embedding the ongoing services there was some tidying up work to be done on the headings Marcive had been unable to change and had notified us of in unrecognised headings reports. The personal names report alone had 250,000 lines and was so big it would not fit into one single excel spreadsheet. We employed a temporary member of staff and asked him to create a separate file of names appearing on this report more than three times. This was on the assumption that we have some unusual material at LSE and that name authority records were less likely to be available from the Library of Congress for names occurring only once in our catalogue. As well as dealing with these multiple occurrences he was also able to work on the unrecognised subject headings

report. We designated reports on corporate names, meeting names and series names as lower priority. Once the temp had completed high priority work we carried out a cost benefit analysis on the merits of completing the outstanding reports. Our sample testing suggested that authority records would not be available for over 90% of the remaining headings (those not already corrected by Marcive) so measurable benefits would be few in relation to the amount of work required in terms of time and cost.

The project could not have been completed without the hard work of those in the Bibliographic Services team who took part in testing data, our senior library assistant who oversaw the work of the temp and contributed in many other ways, and our IT team who persevered with the technological challenges.

We have been delighted with the result of all our hard work. The profile of authority control has been raised within Bibliographic services, combined with our ongoing services from Marcive means we are in a strong position to keep the catalogue in good condition as we move forwards. As a result of the project our catalogue is now a great deal more consistent and has considerably less errors. In a library this size the catalogue is the primary way in which users identify the material we hold and so anything which makes that easier and improves accuracy is surely worth the effort involved.

A fuller write up of the project will be available in a forthcoming issue of **CILIP's Update magazine**.

- based on the presentation given at CIG authority control workshop 2009, London

Introduction and background

The control of series titles as access points in a library catalogue serves as a useful means of collocating related works, similar to the use of uniform titles for works in translation or classics such as those by Dickens and Shakespeare. The principles of series authority control may be set out as follows:

- Ensure that all works in a series are retrievable with a single heading.
- Ensure that the heading (uniform title) for each series is unique.
- Save the catalogue user time and effort by consistently using the heading established for the series.

AACR2 (2002 edition) allows for series titles be given added entries (Rule 21.30L) as well as being part of the descriptive elements of a bibliographic record. Many series could, until recently, be given added entries using the 440 field:

440_0 #aStudies in modern history

If qualifiers needed to be added to distinguish between series with identical titles the 490 and 830 fields could be used.

490_1_ #aStudies in modern history
830_0 #aStudies in modern history
(Oxford University Press)

490_1_ #aStudies in modern history
830_0 #aStudies in modern history
(Yale University Press)

Also, if a series title appearing on a book did not appear to be suitable

for retrieval purposes a uniform title could be created and the 490/830 pair used in the bibliographic record:

490_1_ #aDK eyewitness travel
830_0 #aEyewitness travel guides

Changes to Library of Congress and British Library policies

This situation changed on 1 June 2006 when the Library of Congress decided that it would no longer create and update series authority headings and thus would no longer provide series access points in bibliographic records it created after that date. It would only use the 490 field (first indicator 0 – series not traced to 830 field) for all new LC records. So a series title on a record would only be used for description. Thus the collocation provided by series headings would be lost.

The LC decision was followed 2 years later by the decision by MARBI to make the 440 field redundant in favour of using the 490 and 8XX fields for traced series. The British Library also announced in 2008 that it would change its policy on recording series information in the records it creates. Its practice now is to record the series appearing on books in a 490 field and use the 830 field only when a series title is 'generic' i.e. 2 identical series titles but different issuing bodies e.g. Occasional papers, Research reports.

Implications of these change

The above changes to how series are recorded in MARC21 will be followed by libraries who wish to avoid changing bought-in records from cataloguing agencies but the result will be a lack of collocation

by series in catalogue displays. As added entry fields 440s could be hyperlinked in Web-based catalogues to series authority records, enabling series headings browsing in OPACs.

For example :

Works

128 Teach yourself books
1 Teach yourself history library
7 Teach yourself literature guides
5 Teach yourself revision guides

(Source: Inverclyde Libraries OPAC)

490 fields have no such hyperlinks because they do not serve as access points. Retrieval of series can now only be done by series title keyword so the actual series will no longer display in a browse able index as was possible before. While the 490 field can be used to trace series to the 830 field, this should not be done if the series statement requires no amendments for retrieval or additions such as the issuing body:

490_0_ #aTeach yourself books

Many series in new books, particularly those acquired by public libraries, require no additions to distinguish them from other series. As a possible alternative solution, some library management systems may allow for the 490 field to be suppressed from the OPAC but this may have to be done at a charge from the LMS vendors.

Practical applications of the new rules in libraries

Libraries have reacted in different ways to the LC, BL and MARBI

decisions on series headings. Some continue to use the 440 field for series titles, presumably maintaining their own headings in-house without an external source of authority headings such as LC Authorities to check.

Below are examples of series headings policies in libraries:

Inverclyde Libraries

At Inverclyde we have decided to use the 490 field (first indicator 0) for all fiction series to avoid the extra work of changing records obtained from cataloguing agencies. For non-fiction series we trace titles to the 830 field if they need to be altered to improve retrieval or to distinguish between generic series titles, adding the publisher/issuing body in parentheses.

National Library of Scotland

NLS continues to index 440 fields. If it is necessary to add a series entry to an existing MARC record, then either a 490 0 or a 490 1 and 8XX field will be added. When creating a new MARC record the NLS will use either a 490 0 or 490 1/8XX pair of fields. It uses

authorised series headings whenever they are available. The 490 field is not indexed on Voyager LMS so is of little use to users attempting a series search on the NLS catalogue.

St. Andrews University Library

St. Andrews University Library's policy is to use a 490 1 the series statement should be indexed in its authoritative form i.e. an 830 should be supplied. For new downloaded records cataloguers are required to exercise judgement on how much editing of a series should be done. It still accepts records with 440 fields as well as those with 490/830 pairs.

Conclusion

After following debates on various professional email discussion lists/ blogs it appears that the consensus is to use 490 1 and 830 wherever possible to enable libraries to display series titles as headings in their OPACs and thus continue to collocate related books. The alternative is to rely on series title keyword searching. Since catalogues are designed for the benefit of library customers, cataloguers need to consult them to

see what effect the changes to series authority control policies have had on the retrieval of bibliographic records.

Colin Duncan, Electronic Services Librarian, Inverclyde Council
e: colin.duncan@inverclyde.gov.uk

Post notes of interest, feedback and suggestions for topics to be covered by C&I at the CIG blog:

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



Cataloguers are increasingly moving away from the familiarities of tradition to embrace (or at least, work within) a technology-dominated world of digital records, online cataloguing consortia and new models of resource storage, description and access. Clearly this should not mean leaving behind the skills built up over generations at the portal to a 'library without walls'. It should involve emphasizing them, using them (and adapting them) with open minds, in new contexts. With OPAC records now commonly considered alongside electronic and online resources, at both local and international levels, expertise in organisational systems of all kinds should be productively shared between librarians and computer scientists (who after all, are not such different beasts). These dialogues are beneficial both for information professionals and the users whose needs they serve.

At the same time, the popularity and development of 'social tagging' websites - such as del.icio.us, reddit.com and flickr - encourages members of the public to engage in a form of mass, manual subject indexing. Some are better organised and more conscious of the benefits of tradition and standards than others. The online encyclopedia Wikipedia is often criticised both for its ubiquity in Google search results and for the fact that it 'undermines' peer-review. Often overlooked is the fact that Wikipedia has developed a unique 'folksonomy' - at the same

time as it instantiates clearly-defined, hierarchical organisation schemes which offer multiple entry points and which freshen up the utility of hypertext.

Wikipedia makes strong use of categorisation, classification and metadata elements; some of its content is exportable via the Resource Description Framework (RDF); a number of its projects wrestle with controlled vocabularies, encouraging authors to use 'utility templates' and 'info boxes' with set metadata fields (such as, to list a few examples - 'alternative names' and 'date of birth' for biographical entries, 'geographical co-ordinates' for countries, or 'founder', 'main ecclesiastical polity' and 'congregations' for Christian Denominations). Its editors constantly strive to increase reliability, neutrality and standardisation. In practice, the site has much in common with libraries offering digital content.

If the World Wide Web and 'Web 2' are to be fully engaged with by libraries and treated by them as collaborators rather than competitors (or simply promotional tools), it is worth developing some knowledge of how a major online information provider like *Wikipedia* catalogues its content. This is especially true at a time when internet search engines are the first port of call for the majority of information seekers.

A focus on the definite problems of 'folksonomies' (often confused with 'social tagging') - where they are seen as unregulated, spontaneous and individual-centric enterprises with little terminological control - as potential players in the development of a Semantic Web, has obscured awareness of more positive steps taken by the internet's informal metadata practitioners. In 2005 Deutsche Nationalbibliothek (DNB) contacted the volunteers of the German Wikipedia's 'Personendaten' (Personal data) project group to suggest that the national library's 'Personennamendatei' (PND) name authority record numbers be incorporated into the site's biographical articles. This bold move illustrated the growing relevance of Wikipedia and its commonality with professionals working in the LIS sector.

In another illustration, Wikipedia's special 'Book sources' page locates books using their ISBN and links to the catalogues of libraries, booksellers and publisher's databases as well as providing bibliographical information via OttoBib. It also has a virtual 'reference desk' to answer enquiries, on a topical basis.

How do the challenges facing 'Wikipedians' overlap with those facing traditional cataloguers in an electronic environment? How has Wikipedia approached cataloguing

its content? How has its system evolved? This article presents a brief overview.

Organisation of content and entry points

Wikipedia uses reference lists, indices and content lists to organise its content (which is largely encyclopedic but which also contains almanac type entries). These are:

Overviews of Wikipedia

Overviews survey what is covered or included in an area

List of overviews - key articles organized by subject

List of academic disciplines - Wikipedia arranged like a college course curriculum

Featured content

Representing the best of Wikipedia, featured content undergoes vigorous peer review.

Featured articles - What editors believe to be the best articles in Wikipedia.

Featured pictures - Shocking, impressive, and informative images.

Featured lists - The 'best' lists in Wikipedia.

Featured portals - Portals regarded as being particularly useful, attractive, and well-maintained.

Featured topics - Topics believed to have coverage which is both comprehensive and well-written.

Featured sounds – Sounds deemed beautiful, impressive, and informative.

Lists

Wikipedia has thousands of topic

lists; some are even lists of other lists.

Lists of topics - a comprehensive list of article lists, arranged by subject

Lists of basic topics - links to key topics on major subjects

Category: Lists - a list of lists in the category system, arranged alphabetically

Two of the broadest collections are:

Lists of countries and many lists by country

Lists of people including by nationality and by occupation

Glossaries

List of glossaries - a list of |glossaries arranged by subject

Category: Glossaries - glossaries in the category and subcategories, arranged alphabetically

Portals

Portals are pages that feature selected articles, images, and often news items about the Portal's theme subject. They also include topic lists, category lists, and to-do lists which are used mostly by Wikipedia's editors. Portals can be found at:

List of portals - a list of active portals

Category: Portals - portals in the category and subcategories, arranged alphabetically

Timelines

Timelines are lists of articles organized chronologically. These are the top-level timelines and lists of timelines:

List of timelines - a selection of historical timelines, arranged by subject. More can be found in

Category: Timelines

List of centuries

List of decades

List of historical anniversaries

2009 - major events this year

Portal: Current events - featured current events and related project activities.

Recent deaths - notable recent deaths, by month

Category: Graphical timelines - graphical timelines in the category and subcategories, arranged alphabetically

Indices

Alphabetical Indices

Complete alphabetical index - sorted by the first 2 letters of the title, e.g. "Aa Ab Ac Ad..."

Categorical indices

Wikipedia's main categorical index system is automatically generated from information (category tags) at the bottom of each article. The category system has 3 top-end pages, which are:

Categorical index – an index of major categories, arranged by subject - that section of the page is an exception to the category auto-generation rule, as it is crafted by hand.

Category: Categories - the highest level or "root" category in Wikipedia - its auto-generated entries are listed at the bottom of the page.

Category: Contents - the category equivalent to this page.

Category: Articles - the category in which all article category systems are located.

Wikipedia's other broad categorical indices are:

List of Dewey Decimal classes - lists the top two levels of this library

classification system

Library of Congress Classification

Wikipedia: Outline of Roget's

Thesaurus - articles organised into a system based on six classes, with thousands of branches, following Roget's system

Spoken articles

Category: Spoken articles

Every article has a list at the bottom of all the major categories it belongs to.

Both lists and categories assume a navigational role in Wikipedia, and are continually being revised and updated, providing a loose and varied mesh of entry points into the collection. Wikipedia attempts to reflect different world views by offering multiple categorisations and classification schemes. It also uses these schemes to encourage growth and development by, for example, using red text to show that an article does not yet exist but that one would be useful. Ranganathan's notion of the "library as a growing organism" is played out nicely here. To this we might add that Wikipedia strives toward a certain transparency of process. Policies, guidelines and tutorials introduce users to the site and explain how they can add and edit articles.

As with libraries there is, on one hand, a well thought out, underlying organisation which helps Wikipedia's cataloguers (i.e. its authors and editors) construct and standardise articles, formalise how they are dealt with and determine into which categories they are placed and, on the other, the simplified, intuitive, organisational layer presented to users as they are guided towards content. Terms (most often names and places) are

disambiguated where necessary and, if a simple 'redirect' page is not appropriate, hyperlinks to probable relevant articles will be presented.

Back in November 2001 the English language Wikipedia had only a very loose set of suggested categorisation schemes, providing an index of all pages on the site (just over 19,000 articles, compared with the 2,280,000+ today). With a talk page to discuss a variety of potential schemes, it allowed users to organise pages according to Dewey Decimal, LCSH, a complete list of Encyclopedia topics, a three-part category scheme attributed to Thomas Jefferson (inspired by Francis Bacon) based around 'Memory, Reason and Imagination' and a 'This Day in History' feature. There was a suggestion that the UseNet dot classification scheme be developed for Wikipedia.

At this stage, the present MediaWiki software (which supports rich content generated through specialized syntax to render for example mathematical formulae, graphical timelines over mathematical plotting, musical scores and Egyptian hieroglyphs) had not been developed. The site relied on the older page-based UseModWiki engine.

As Wikipedia grew, the need for sophisticated software capable of handling its content and an increased user base were required. Similarly the need for a straight-forward and consistent classification and navigation scheme, familiar across the various Wikipedia language sites, was apparent. The new model of information seeking created by the internet's hypertext linking system, and developed/extended by Wiki technologies, began to suggest a

useful strategy for Wikipedia.

These quotes from Charles R. Matthews, co-author with librarian Phoebe Ayers, of the published '*How Wikipedia Works*' (No Starch Press, publication date: June 2008) and a mathematician who sits on Wikipedia's Arbitration Board, are enlightening:

From my point of view, much of the criticism aimed at Wikipedia, and perhaps some of the praise it gets, is misconceived because it takes the "article" as the unit.

As in a search process where the reader:

- 1. Searches for an article, having a specific topic or title in mind;*
- 2. Gets to the article by a rational search process;*
- 3. Reads down the whole article;*
- 4. Decides what to do next.*

Wikipedia shows its strengths in other ways, where readers:

- 1. Have a need for background or just a vague idea ("I need to know more about heart disease")*
- 2. Arrive by surfing and other forms of navigation, for example by moving up and down within*

Wikipedia's category system sampling articles;

- 3. Skip out of articles as soon as they find a promising link (centrifugal motion);*
- 4. Work out as they go along where they really need to be informed (centripetal motion).*

How you behave at point 3 tends to differentiate adult learners (who may put emphasis on thorough assimilation as the way forward), and the Internet generation with an attitude of speedily cutting losses and backtracking any time you are moving off-topic.

Although Wikipedia attempts to democratically develop its own 'folksonomy' (where categorisation and keywords are developed by users and approved by community consensus) it is genuinely collaborative rather than simply collectivist as some 'social networking' sites are. It includes formal systems such as DDC, LCSH and Roget's Thesaurus in its scheme (although admittedly these are not fully developed). At the same time major, separate projects such as Semantic MediaWiki work alongside Wikipedia to allow the incorporation of Wiki records into other sites through the encoding of semantic data and RDF exporting.

The main difference between libraries and Wikipedia in the present information landscape is not a philosophical but a legal one: libraries do not, on the whole, create content; they do not have the luxury to work under a 'copy left' system of licensing; however, the idea of enabling a free, open social dialogue is surely an ideal with which many librarians would find no fault.

There are many entry points into the subject of Libraries/Librarians on Wikipedia.

Wikipedia's Library and Information Science Portal is impressive and crying out for contributions:

http://en.wikipedia.org/wiki/Portal:Library_and_information_science

This portal was recently starred as a 'featured portal' and contains pictures, quotes, topics, news articles and definitions.

There is also the Topic list:

http://en.wikipedia.org/wiki/List_of_basic_library_and_information_science_topics

where there are for example, articles on Library 2.0, Social Cataloguing and the History of Library and Information Science.

A project established by librarian Michael Sauers and others is designed to introduce Librarians to Wikipedia and vice versa.

http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Librarians

It opens with the line: "We librarians

flatter ourselves that we know a thing or two about organizing information. It's time we stepped up and contributed to Wikipedia: not just to its content but to its structures and technologies."

Of course there are problems with Wikipedia - vandalism, articles which lack proper citation, a lack in some places of clearly-defined enforceable standards, issues of anonymity and un-verifiability. But these are problems which Wikipedia users actively work to correct, something many in the online community do not even attempt, preferring the anarchic approach which cataloguers dread; librarians and cataloguers may have much to contribute, and much common cause with the Wikipedia community.

Kathleen Menzies, Researcher, Centre for Digital Library Research (CDLR) University of Strathclyde e: klmenzies@cis.strath.ac.uk

Extras

CIG is pleased to have been able to make a donation of GBP100.00 to both Book Aid International and the CILIP Benevolent Fund. The group has made such donations in the past when its funds allow. Wendy Taylor, CIG Treasurer, is in the process of putting together the annual accounts and it looks like CIG has had a successful year. CIG receives its income from a capitation payment for each member by CILIP. It really does make a difference what CILIP special interest groups people subscribe to, so please encourage colleagues to **join CIG**.

CIG supplements this income with money made from events and the committee is in the process of planning events for 2010, including the **CIG conference**.

If you have any ideas for events please let us know by contacting a member of the committee, all suggestions welcome!

It is a near certainty that the majority of UK libraries make use of, and benefit extensively from, the work that goes into creating and maintaining the LC/NACO Authority File; yet neither the NACO programme (the name authority component of the Program for Cooperative Cataloging) nor its data enjoy a high profile on this side of the Atlantic. Despite the many benefits that accrue to UK libraries, many remain unaware of its existence. Its influence is, however, pervasive, and its importance in the current environment hard to over-emphasise

The work that the NACO partners put into creating authority data and establishing authorised forms of headings for persons, corporate/conference names, jurisdictions, works and expressions (name/title, title and series) has a ripple effect on the whole of the bibliographic universe – or at least on that part of it which is focused on library catalogues and, in particular, the Anglo-American cataloguing tradition. A library that sources most of its catalogue records from the British Library, the Library of Congress, OCLC, RLUK, major vendors and other commercial services is likely to be obtaining data in which most of the name headings conform to authorised forms found in NACO authority records. If you send out your library catalogue for “cleaning up”, then the chances are again that they will have been matched

against the LC/NACO file.

As its simplest, NACO is a community effort that aims to provide most of the authority records that will be needed to support authority control in library systems. Currently it’s supplied only in MARC 21, but its release as part of the Library of Congress “Authorities and Vocabularies” service is expected shortly, and this will extend the range of formats in which the LC/NACO file is made available.

The origins of the file itself can be traced back to the development of the MARC Authority format in 1976. Quickly, the Library of Congress started creating machine-readable authority records. In October 1977 the US Government Printing Office joined with LC in sharing authority work. Texas State Library joined in during 1979, and the start of the next decade saw what was to become an explosion in membership, with a total of ten members by 1980, 24 by 1982, 37 by 1985, 55 by 1992, and so on. In Fiscal Year 2008 no fewer than 379 institutions contributed, of whom 51 created more than 1000 records.

Participants come from a wide range of backgrounds – everything from national libraries, the libraries of other national institutions (such as the US Army), through various tiers of education (albeit with higher education dominating), public, state, etc., libraries, organisations such as OCLC, publishers, vendors, and more. (This list should

not be read as implying any particular hierarchy of importance!) Many of the smaller contributors provide vital specialist input; often these are grouped into Funnel Projects, both to aid administration and to foster a sense of community within a particular group, whether it be subject- or language-based, or simply organised around a specific geographic area.

For all but the most specialised collections, the LC/NACO file should be able to meet the overwhelming majority of the needs of staff, end users and library systems. One of the difficulties in determining its effectiveness, though, is that there is no formal list of expectations against which to measure “success”; nor is there even an agreed definition of “authority control”.

Membership of NACO brings obligations and expectations. Contributors are obliged to follow standards, commit staff, undergo training, contribute through a utility, meet minimum contribution levels, and be capable of achieving total independence from their assigned mentor within 12 months. The standards are not just the obvious ones (AACR2, MARC 21), but also include those less often encountered in UK cataloguing departments (LC Rule Interpretations, ALA-LC Romanization Tables). Indeed, it can sometimes seem as though there’s too much documentation to

consider – or at least that there are too many places to look!

The full training programme takes 5 days, although it may be possible to compress this. More problematically, there are currently no NACO trainers based in the UK. The PCC is actively looking at online training to substitute for (or to supplement) more traditional means. Clearly, though, a would-be member is required to make a significant investment of time and effort.

There are currently 8 UK members plus one in the Irish Republic (Trinity College Dublin); two of these (Oxford's Bodleian Library and Cambridge University Library) also contribute series authority records (an activity which requires additional training). Since all of the UK legal deposit libraries are members, virtually all headings in non-CIP British National Bibliography records conform to authorised LC/NACO forms. Over 25% of the non-LC additions to the LC/NACO file come from outside the US – and of that number around half are contributed by our own British Library, by a long way the most productive contributor to the file apart from LC. The total size of the file now exceeds 7 million records.

Members not only establish new headings, but also undertake selective maintenance on existing authority records. The relative antiquity (in machine-readable terms, that is!) of some of the data – the earliest records predate AACR2, for example – means that there is rarely any shortage of

“clean up” work required, just as most libraries are doubtless themselves aware that work on maintaining their own databases is a task that can never said to be “completed”. But the NACO programme aims to limit changes to authorised forms of heading to those that staff or users might consider “important”; change for the sake of change is definitely not part of the philosophy!

Some might ask why any library might want to become a member. Partly it's a recognition that the only way of reducing cataloguing costs is through cooperation, locally, nationally and internationally. This is something that has been recognised in the context of bibliographic data for some decades; but the same principle applies equally to authority data. The more there are contributing to its creation, the less the overall cost for any one of us. And, as indicated above, there are many who benefit from this work, even if they don't even have authority records in their library systems. Giving something back to the community is a natural extension.

Membership quickly pays dividends; and, whilst gritting of teeth is occasionally required, it is a good discipline (cataloguers think that extra bit more about what they're actually doing) as well as a rewarding activity (there's more than enough satisfaction in the challenges to overcome the occasional frustration, and a genuine sense of achievement). Plus contributing “your” form of heading has practical benefits

when it comes to maintaining the integrity of your local catalogue.

However, nothing in life is perfect, and many accept that the community – by which I mean all of us! – need to be looking at even more powerful and effective solutions. These will more than likely not depend on an “enclosed” community of contributors, and will surely take advantage of technological solutions that could not have been dreamed of when LC first began to create authority records back in 1977. In the meantime, though, the achievements over the years have been tremendous and are a firm foundation to build upon for that future.

Further information

Program for Cooperative

Cataloging

<http://www.loc.gov/catdir/pcc/>

NACO

<http://www.loc.gov/catdir/pcc/naco/naco.html>

Library of Congress Authorities
(free access to LC/NACO authority records)

<http://authorities.loc.gov/>

Library of Congress Authorities and
Vocabularies service

<http://id.loc.gov/authorities/>

LC Cataloging Distribution Service

<http://www.loc.gov/cds/>

*- based on a presentation given at
CIG authority control workshop
2009, London*

Hugh Taylor, Head of Collection
Development and Description,
Cambridge University Library e:
jrht3@cam.ac.uk

The main focus of this essay lies in highlighting the strengths and weaknesses of pre-coordinate and post-coordinate retrieval systems, providing 'real-life' examples of each system, and explaining some of the ways in which these two systems can be used and exploited. Thus, it is crucial that the broader system that these two systems are part of is discussed with the aim of supplying clear definitions, and an understanding of the two systems under review.

Indexing is defined as the process of representing the subject of a document. Jones et al postulate that indexing is twofold: *"to assist in the retrieval of relevant documents in answer to a query; and, to suppress the retrieval of non-relevant documents"*¹.

Indexing in itself is generally an intellectual process involving a great amount of subjective judgement. As a result of this human input, it is extremely costly to perform on anything other than a very small scale. Furthermore, indexing can be responsible for a significant proportion of information retrieval failures; hence, it is a major concern in examining information retrieval systems (IRS).

Jones et al cite that *"indexing is a multi-stage process in which the success of each stage is dependent upon the preceding stages"*². This process was subdivided into the four stages: content analysis, concept selection, concept translation, and term combination. These processes are the same for both pre-coordinate and post-coordinate systems. Content analysis concerns itself with the subject or subjects of a document.

The concept of 'aboutness', which is the source of considerable inconsistency in indexing and central to the entire process, is recognised at this stage. Concept selection entails establishing aspects of the subject matter of a document that are of interest to the users of the IRS. In essence, after it has been determined what the document is about (content analysis), it is time to take into account the requirements of the users of the IRS. Further policy decisions at this stage is concerned with exhaustivity, which is the depth of indexing, that is, how many of the concepts mentioned in the document are to be indexed individually; and specificity, which relates *"to the exactness with which the index terms match the concepts in the document"*⁴. The consensus is that specific terms are to be preferred to more general terms, for example: 'Chihuahua' rather than 'dog'. This practice avoids problems associated with retrieving large numbers of documents when indexing has been carried out using general terms. Concept translation is concerned with providing *'the appropriate terminology for representing the document subject in the IRS'*⁴. Thus, concept translation is the process of turning the indexer's ideas about the subject or subjects of a document into index terms. However, before concept translation is delved into, decisions regarding the source and nature of the vocabulary have to be made. Source of terms in indexing are considered to be derived or assigned. Derived terms are those extracted directly from the text of the document. This involves the use of an uncontrolled vocabulary, which is, using the author's own words or 'natural language' as

index terms. On the other hand, assigned terms are not restricted to terms extracted from the document but are terms assigned at the indexer's discretion which are representation of the document's content. This ensures that concepts implied by the author that are not mentioned explicitly in the text are indexed. This requires greater judgement on the part of the indexer and tends to be associated with the use of index terms drawn from a controlled vocabulary. A controlled vocabulary is defined as *"an artificial language designed to overcome the variability and ambiguity inherent in natural language"*⁵. Basically, it is an authority list. It serves to improve the likelihood of a match between the terms used by the indexers and searchers. Both pre-coordinate and post-coordinate systems employ controlled vocabularies. A controlled vocabulary has the following advantages:

- i. It lists the range of terms that must be used in indexing and searching, and refers the users from the non-preferred to the preferred terms. For instance, if 'cows' is the preferred term, indexers and users would be directed from 'cattle' to 'cows'.
- ii. It makes ready-made compound terms available. These terms comprise two or more elements or facets, for example: "butter production or 637.2"⁶.
- iii. It distinguishes between homographs, for example, pitch (Bitumen), pitch (football), pitch (music), and pitch (slope).
- iv. It links terms together that are

in some way related, for example, 'cows' and 'milk': this should aid both indexers and searchers in term selection by suggesting more suitable terms.

Term combination is the final stage of the indexing process, and it is concerned with decisions relating to the order of the index terms. This is determined by the type of indexing system utilised. Indexing systems are divided into two categories: pre-coordinate and post-coordinate systems. Jones et al posit that the distinction between these approaches occurs on how they handle compound subjects.

"Pre-coordination is the combination of index terms at the indexing stage"⁷. The indexer constructs a heading comprising as many terms as are required to summarise as much of the subject content of the document as the indexing system permits. Essentially, the subject access points comprise strings in their entirety. Rowley and Farrow suggest that this approach reflects and systematises our natural tendency to think of subjects as title-like phrases. For example: 'skin diseases in dogs'. Rowley and Farrow supplies the Library of Congress Subject Headings (LCSH) as an example of a pre-coordinate indexing system. Pre-coordination is used with shelf-arrangement, library catalogues and bibliographies, classification schemes, alphabetical subject catalogue, feature headings, card catalogue, book or printed forms, and printed indexes as according to Foskett. Pre-coordinate indexes are critical to manual searching, and traditionally two forms of manually searched indexes have used pre-coordination: dictionary and classified indexes as

suggested by Rowley and Farrow. The simplest way to arrange terms is alphabetically as in a dictionary; nonetheless an alphabetical arrangement "*cannot show any kind of relationship except bringing together words which have the same stem, which at best can only cater for a very small part of the problem*"⁸. Foskett states that the answer to this problem is to insert linkages, usually called cross-references. These serve to bring semantic relationships to the attention of the user. A cross reference is made from the non-preferred term to 'USE' the preferred term, for example: footpaths USE trails. The second method of arrangement shows relationships by juxtaposition, which is grouping related concepts systematically to form a classification scheme. The arrangement highlights hierarchical relationships as well as coordinate relationships. This eliminates a substantial part of the cross reference structure required by an alphabetical arrangement. Books on the shelves of a library is normally arranged in this order to assist users in finding all the books they are interested in shelved in the same area. However, this system forced the introduction of notation or code vocabulary to show order and enable users to find concepts among the systematic arrangement. Hence, terms have to be searched in the entry vocabulary which will direct the user to the code used to denote them, for example: electronic – 621.381 DDC. (Foskett, p.93).

One of the strengths of a pre-coordinate system is that it allows for indexes that are familiar to users since they present a more or less complete statement of the subject. Pre-coordination is avail-

able in well tried, 'standard' systems, such as LCC, DDC and LCSH in MARC records. The alphabetical order facilitates usage with little or no training. Rowley and Farrow cite that pre-coordination is the most powerful device for improving the precision of a search – "*far more precise than the crude 'AND' of Boolean searching*"⁹.

Another strength of pre-coordination is that of citation order, which is one of two mechanisms employed to establish the context of a term. Citation order is the fixed "*predetermined order in which subject concepts are to be arranged when the indexer assembles them in strings*"¹⁰. Citation order traditionally has always been based on significance order. It is applied to achieve the desired consistency of indexing different documents and "*to place documents and document records at a useful point in the overall sequence*"¹¹. Furthermore, the strings of concepts assembled according to the citation order serve as mini abstracts to documents. This facilitates browsing for users and highlights the relevance of the document. Bodoff and Kambil mention that the enhanced precision and recall benefits of pre-coordination result from "*the standardisation of term orderings and from the selection of intelligent term orderings*"¹². However, a fixed citation order promotes complications in collocation and searching since only the first term in the index string forms an access point, while documents on concepts occurring later in the string will be scattered, thus promoting difficulty in retrieval.

To offset this problem of scatter, a number of possibilities exist, but

according to Jones et al, none is without problems. Permutation, which is the method of “*manipulating the strings so that every index term forms an access point and every term combination is catered for*”¹³, is impractical since it generates an enormous number of entries. For instance, the three terms ‘ABC’ could be permuted to give six strings; however, seven terms give 5040 strings. This illustrates the number of entries that are derived from permutation. Furthermore, producing this index would be too costly, and the index would be too huge to be used effectively. Moreover, not all the combinations constructed by this type of mechanistic approach would be useful. Another possibility can be rotated indexing, which is a more promising approach. The purpose is to allow each term to form an access point but “*the relative position of the terms in the string remains unchanged*”¹⁴. Although this approach is more flexible than a single string and avoids wasteful duplication, the main problem arises when users require a term combination which is not provided by the index.

Another weakness of pre-coordinate systems is that they can be excessively complex for the searcher, who has to anticipate both the correct terms and the correct combination of terms under which to search. Pre-coordinate indexes are only effective at summarisation level because large numbers of terms in a string become very difficult to handle. Therefore, they are not well suited for exhaustive (depth) indexing. Moreover, indexing in controlled language systems is slow and costly. However, this is not the case with title-derived

KWIC and KWOC indexes as according to Rowley and Farrow.

Bodoff and Kambil state that an often overlooked advantage of pre-coordination is that standardisation of order enhances precision when the same terms can express different ideas through different syntaxes. For instance, ‘oil crises [in America] result in U.S. invasion [on Iraq]’ is different form ‘U.S. invasion result in oil crises’ (although we know this was the case). In both instances, the same terms are related by a different *cause and effect* relationship. Pre-coordination, theoretically can define the standardised ordering of concepts when two terms are linked in such a semantic relationship. When this occurs, this *effect* term must always precede the *cause* term. This ensures that the two aforementioned examples have entirely different meanings, and appropriately assigns documents to distant parts of the card catalogue to avoid confusion on the subject. In post-coordination, however, the unordered list of query terms ‘oil crises, U.S. invasion’ matches documents on both subjects. Bodoff and Kambil further emphasise that intelligent term orderings enhances precision since similar documents are grouped together while un-grouping dissimilar ones.

This eliminates improper partial matches by clearly defining the context of each heading term. Partial matches in pre-coordination result when the searcher has guessed the first term of a subject heading in the correct order but has failed to guess all of the terms.

The second mechanism utilised to establish the context for a term in pre-coordinate schemes is (inverted) term phrasing.

According to Bodoff and Kambil, the LCSH heading ‘Art, Asian’ is the term phrase used to ensure that the term ‘Asian’ matches only in the context of Art. With a compound subject, the indexer curb spurious partial matches by either forming a phrase or using citation order between terms.

The attempt to combat problems that relate to efforts to impose a structure when forming a subject index string, which is associated with the pre-coordinate system, saw the development of the post-coordinate system from 1940s onwards. Nonetheless, Foskett emphasises that the post-coordinate system was established to clarify the difference while confirming the essential similarities of the pre-coordinate system. Hence, the post-coordinate system is used “*to denote indexing in which concepts were coordinated at the time of searching*”¹⁵. This method eliminates the need for citation order, facilitating faster and cheaper indexing but loses “*the element of meaning provided by the syntax of the string of pre-coordinated terms*”¹⁶. The index terms may be either single words or compound terms consisting of two or more words. However, these compound terms represent single concepts as in ‘dry cleaning’. This system is referred to as thesauri. With post-coordination, the user does not have to accept the term combinations provided in advance by the indexer but can assemble terms at the search stage in any order as the need arises. Since terms are freely coordinated by the user at the output stage, it is much easier to search for multi-concept subjects. For instance, let us suppose the indexer used the following terms in classifying a document: Socio-linguistics,

Jamaica, Creole Languages, Jamaican Creole, and Post-Creole Continuum. This is exactly where the indexer's work ends as there is no need to impose citation order. The searcher can retrieve the document by using one or a combination of index terms.

Jones et al mention that the first post-coordinate systems were designed for manual operation using cards. The approaches employed were Uniterm Cards, Peek-a-boo cards, and edge notched cards. Searching a manual post-coordinate index required skill and was time-consuming. Post-coordinate systems became important with the advent of computerised IRS.

Like, pre-coordinate systems, the strengths of post-coordinate systems are counterbalanced by a number of weaknesses. The most obvious is the problem of false coordination of search terms. With false drops, the terms required are actually found in the document but the document is not on the required subject. This problem is due to the lack of syntax. False coordination promotes information overload, especially among computerised systems. Bodoff and Kambil mention that *out-of-context* matches are the primary sources of false drops, and stress that understanding the types of out-of-context matches is critical to improving search methods. The five types of out-of-context matches are prominent in computerised IRS.

i. Ordered relationship among terms

Polysemy - This occurs when a term in the document index differs in meaning from the same term in the query.

Out of phrase terms - An example of this is an individual term occurring within a phrase in the document heading, and in a very different phrase or as an individual term in the query or vice versa.

Exhaustivity: Secondary Topic Keywords - In this sense, "*the matching terms are assigned unwarranted prominence by an algorithm which does not understand the primary topic or context of the document*"¹⁷

Depth: Non-categorical terms - The false drop occurs when the narrower term matches out of the context its intended broader category.

These are all evident in the World-Wide-Wed phenomenon, the utilisation of search engines that return thousands or sometimes millions of irrelevant documents in response to a query.

Due to the multi-dimensional nature of the post-coordinate system, the need for users to learn formal rules to construct searches such as citation order and vocabulary standards is reduced. This method accommodates different types of search patterns to be facilitated through the use of Boolean operators, but 'formulating Boolean searches and the protocols of machine searching can be complicated, even in menu-driven systems.'

Post-coordinate permits indexing to any level of exhaustivity, and a higher recall than pre-coordination since all documents with the requested coordinates are retrieved. Nonetheless, a higher recall diminishes precision. Hence, searchers are faced with the onerous task of wading through all of the retrieved items to locate

relevant documents. From experience, a large number of relevant documents go unnoticed as the searcher may peruse the first few documents only, as time-constraints may hinder the viewing of all information retrieved. Post-coordinate systems are not browsable but most effective when users specify their search needs.

In summary, the pre-coordinate approach arranges items in a logical order, which is linear and mono-dimensional as exemplified by goods on supermarket shelves and correspondence files in filing cabinets. Printed information sources, because of their linear nature, are usually pre-coordinate. However, this approach may be also implemented in an electronic environment as seen in a menu of choices presented by a computer system which asks users to make selections on a hierarchical basis. This method is quite popular in videotext systems such as CEEFAX, ORACLE and MINITEL. It may be also found in OPACs, gophers, and internet search engines such as Yahoo! and Excite. Pre-coordination is useful for general searchers, who may be vague about their information needs and who may be making a general search rather than comprehensive answers to complex questions. Different pre-coordinate arrangements may be employed within one collection, for instance, entries in a printed directory may be presented in one sequence (alphabetical by title), with additional indexes giving different sequences (author, subject, and place). According to Jones et al. pre-coordinate indexing systems are best used:

- for 'mark and park', which is providing a shelf number, where a single place is required as the

physical location. An example of this was mentioned with 'butter production' which will be shelved at 637.2.

- where depth indexing is not required.

The post-coordinate approach does not attempt to utilise the logical sequence since there is an independent retrieval mechanism which allows for the combination of multiple search terms. Browsing is not catered for unless some complementary retrieval tool is provided as constructed by Burke. Document location is irrelevant as the document is retrieved by linking various descriptive components together. Access points provide the basis for retrieval. These may be subject, date, or some other characteristic. The document may be arranged in some way which may be convenient from an administrative viewpoint, but does not claim to be logical for the searcher. Post-coordination may be applied to both the original items and their surrogates. An example of this is a library OPAC.

Computers provide an efficient means of carrying out coordination. However, post-coordination may be implemented in any of the following systems – manual, mechanised, and computerised. Post-coordinate indexing systems are best used as according to Jones et al -

- where depth indexing is needed, since it is possible to handle large numbers of index terms when they are all indexed individually.
- for highly specific search need.

Bodoff and Kambil propose a new approach called *Partial Coordination*, that combines the

strengths of the two subject approaches – pre-coordination as in traditional card catalogues, and post-coordination as in a computerised keyword search. This method is aimed at a greater precision for a given level of recall. In addition, it is supposed to alleviate the problems associate with all out-of-context matches except polysemy. Nonetheless, research reveals that this notion of partial coordination is not popular among writers, so it is safe to assume that its popularity has not spread.

In conclusion, the theoretical distinction between pre-coordination and post-coordination may seem clear cut; however, in practice this is not the case as there is some grey area since the two approaches borrow from each other in an attempt to overcome their weaknesses. For example, post-coordination is supposed to restrict itself to single terms in theory but make use of compound terms to enhance precision. Let us consider the concept 'group dynamics' – if these two terms were indexed individually a search would retrieved many irrelevant documents containing one or both of the terms. Hence, the practice is not so clear cut but a continuum exists with pre-coordination at one end of the pole and post-coordination at the other.

References

1. Jones et al. Jones, Philip H., Lucy Tedd and Alan Wheatley. *ILM 7820 Principles for Information Retrieval*. Aberystwyth: The Open Learning Unit, September 2000. Unit 3, p 2.
2. Jones et al. Unit 3, p 3.
3. Jones et al. Unit 3, p 8.
4. Jones et al. Unit 3, p 9.
5. Jones et al. Unit 3, p 10.
6. Jones et al. Unit 3, p 11.
7. Rowley, Jennifer and John Farrow. *Organizing Knowledge: An Introduction to Managing*

- Access to Information*. 3rd Ed. Aldershot: Gower, 2000. p 166
8. Foskett, A. C. *The Subject Approach to Information*. 5th Ed. London: Library Association, 1996. p 85
9. Rowley and Farrow. p 167
10. Jones et al. Unit 3, p 17
11. Lowe and Wheatley Glossary
12. Bodoff, David and Ajit Kambil. "Partial Coordination. I. The Best of Pre-Coordination and Post Coordination." *Journal of American Society for Information Science*. 49.14 (1998): 1251-1269. p 1255.
13. Jones et al. Unit 3, p 18.
14. Jones et al. Unit 3, p 19.
15. Foskett, A. C. p 97.
16. Foskett, p 102.
17. Bodoff and Kambil. p 1259.
18. Rowley and Farrow, p 173

Bibliography

Aluri, Rao, D. Alasdair Kemp and John J. Boll. *Subject Analysis in Online Catalogs*. Colorado: Libraries Unlimited, Inc. 1991

Burke, Mary A. *Organization of Multimedia Resources: Principles and Practice of Information Retrieval*. Aldershot: Gower, 1999.

Cleveland, Donald B. and Ana D. Cleveland. *Introduction to Indexing and Abstracting*. 3rd Ed. Colorado: Libraries Unlimited Inc., 2001

Fugmann, Robert. *Subject Analysis and Indexing: Theoretical Foundation and Practical Advice*. Frankfurt: Indeks Verlag, 1993.

Gredley Ellen and Alan Hopkinson. *Exchanging Bibliographic Data*. London: Library Association Publishing Ltd. 1990

Large, Andrew, Lucy A. Tedd and R. J. Hartley. *Information Seeking in the Online Age: Principles and Practice*. London: Bowker Saur, 1999.

Lancaster, F. W. *Indexing and Abstracting in Theory and Practice*. 2nd Ed. London: Library Association, 1998.

Jennine Knight, Librarian II
University of the West Indies e:
jennine.knight@cavehill.uwi.edu

In this article I want to discuss the future of authority control. Authority control is expensive and is only applied to a relatively small number of resources. While this was justifiable and useful in the context of an institutional catalogue, the value of the investment is more questionable in the context of the hybrid library or the web. If the scope of authority control has to be extended to a much wider range of resources, how can we achieve this at an acceptable cost? How might our processes and data be more scaled? To what extent might automation of the authority control process be possible? This paper evaluates the potential for new models, standards and projects to address these challenges.

The catalogue is an instrument to save the time of the information seeker. Authority control is one of the syndetic structures which turn a list of books into a catalogue. The Paris Principles instructed that, "The catalogue should be an efficient instrument for ascertaining...which works by a particular author are in the library." These principles were updated earlier this year following a lengthy international consultation.

The catalogue should be an effective and efficient instrument that enables a user ...to find sets of resources representing all resources associated with a given person, family, or corporate body
Statement of International Cataloguing Principles, 2009

The *Statement of international cataloguing principles* reiterates the principle of comprehensiveness, but recognising that the computer is not bound by the same physical constraints as the card catalogue, extends the requirement to include any person, family or corporate body associated with a resource. From the perspective of the user this is absolutely right, but, I wonder how many of our catalogues would satisfy this criterion?

In practice we make choices regarding authority control. The choices vary between institutions. The British Library is probably typical in that authority control is disproportionately allocated to printed books and printed music. There is no formal authority control of article level literature and coverage of non-print resources is inconsistent. Inconsistencies also persist between current and retrospective practices, despite much hard work by authority control staff over recent years, to reconcile headings in the integrated catalogue with the LC/NACO Authority File.

As a recent survey by JISC confirms, printed books are not (and have not for some time) been the preferred means by which academics and researchers publish their work. Even in the humanities, journal articles are a key measure of academic productivity.

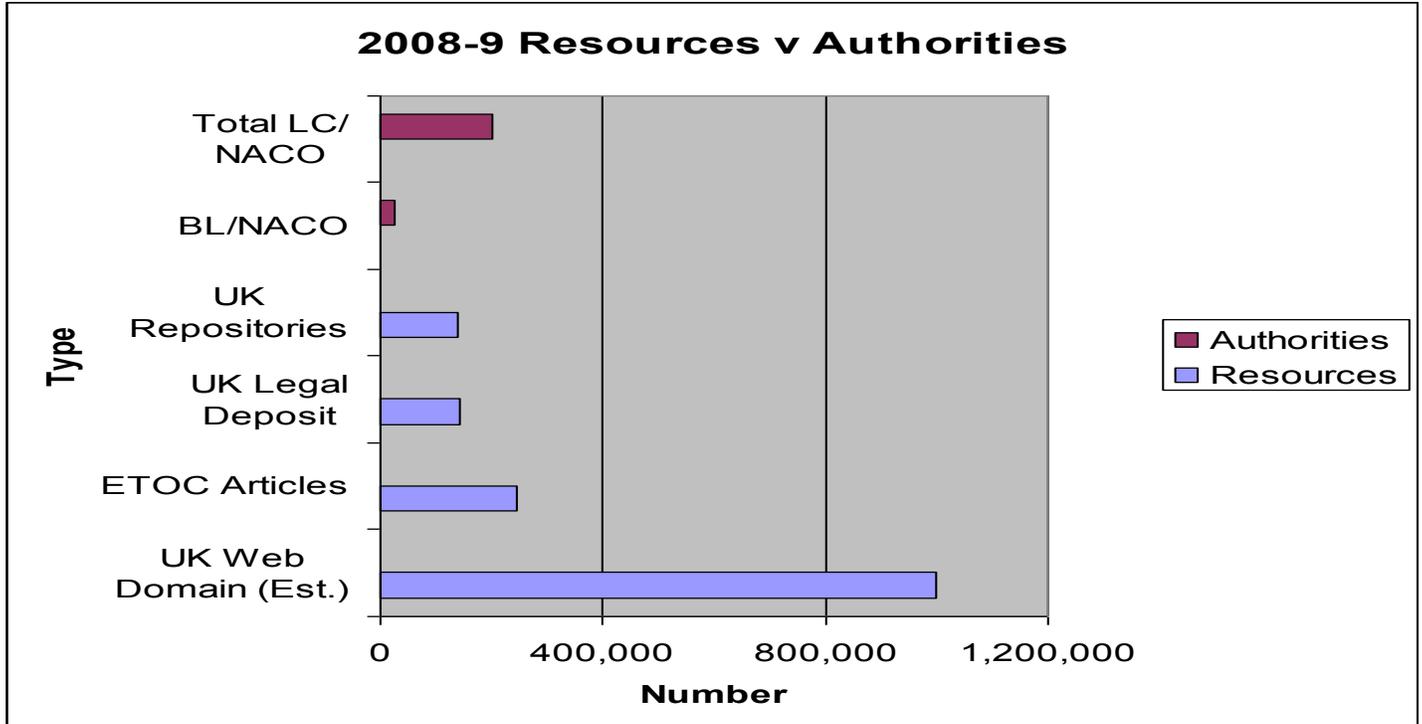
The only major exceptions to the dominance of the journal article is

the continuing high status attached to monographs and edited volumes in the humanities, and to practice-based outputs in the arts. Yet even in the humanities, journal articles are now by far the largest publication format by volume.

In certain fields, data sets, preprints, conference proceedings and web resources are the currency of research.

In computer science, for example, the pace of change means that conferences are particularly important, and these may attract higher prestige than journal articles...Repositories have achieved less traction in the humanities and social sciences than in many science and engineering subjects.

The uncomfortable conclusion is that delivery falls short of aims and ambitions, as expressed through the International Statement of Cataloguing Principles. Names associated with resources at levels of granularity lower than the book or resources that are published or distributed through repositories are not effectively controlled, therefore I can't go onto Google and find all the resources associated with Mao Tse -tung / Mao Zedong /毛澤東, irrespective of whether they are held by a library, archive or museum, or part of the full text or metadata. Even if this were technically possible, I would need to conduct multiple searches to encompass the variant forms of name employed according to sectoral, institutional or national practices.



Whatever the reasons for the constraints, the consequence from the user perspective is that I don't get what I want. Is authority control of printed books adequate? When the web is breaking down physical barriers to access can we afford to neglect the semantic obstacles littering the information landscape? Will administrators think that partial authority control is worth paying for?

What is the size of the problem? In 2008-9 approximately 1.5 million new resources were "published" in the UK. The graph illustrated the breakdown by type of material. This includes an estimate of the number of new websites in the UK domain; the number of articles added to the British Library's Electronic Table of Contents; the number of new titles received through legal deposit and the number of resources deposited in

repositories. It is reasonable to assume that each of these resources must have at least one name associated with it. Not all of these will be new names of course and more research is needed for certainty in this area. The graph also shows that the total number of new names added to the LC/NACO authority file over the same year, was just over 200,000. This represents a daunting shortfall.

On receipt of a book, cataloguers will determine whether the creator, contributor or other names associated with the book have already been added to the authority file. Is the A. Rose PhD on the title page, the same person as Dr. Alex Rose.? The process by which cataloguers do this is a combination of information gathering and matching. Cataloguers find as much information about two identifies as

they can and check for points of coincidence. Cataloguers may contact publishers or authors directly to substantiate information when all other sources fail. Current workflows are fundamentally based on human review and is therefore inherently expensive and do not scale well.

An alternative approach would be to capture authoritative information before a resource is published, so that authoritative information is available throughout the value chain. This means involving publishers, creators, researchers and librarians. The realities of (mis)identification on the web mean that benefits to other sectors and individuals are much more evident than has previously been the case.

The human factor can be supplemented by automation. Automated or semi-automated

approaches for derived cataloguing are commonplace, but less well developed for authority control. Automation would be facilitated by unique, persistent identifiers for persons, families or corporate bodies, just as the ISBN aids bibliographic record matching. The International Standard Name Identifier (ISNI), developed by libraries, trade and rights management organisations, is currently a draft international standard. As well as being an identifier, the ISNI also defines supporting metadata necessary for disambiguation and assignment.

Current authority control practice, as derived from AACR2, is that disambiguation, not identification, is the governing principle. This means that a heading can be established so long as it can be distinguished from all other headings. The consequence of this is that data elements may not be explicitly specified. For example, if not needed to disambiguate a name, the date of birth or an unused forename, may be recorded as a note; even when recorded in the heading, the data is part of a complex string, not a distinct well defined attribute. Information buried in strings or notes, may be accessible to cataloguers, but not to machines. It is also questionable how much assistance it is to an end user to distinguish between A. Rose, born in 1935 from A. Rose who died in 1970. It would be much more useful to the user to know that A. Rose is a woman, who has worked at the University of Somewhere

and has written papers with A.N. Other. To answer these questions, the focus of authority work has to shift to identification.

The Functional Requirements for Authority Data (FRAD) is an extension of the FRBR model to tasks related to authority control. The FRAD model defines entities, attributes and relationships needed to support authority control tasks, including user tasks, such as FIND and IDENTIFY. FRAD supports different implementations of authority data. It looks ahead to an object oriented approach in which attributes are recorded to create an entity record that machines can process. FRAD also defines relationships between entities. FRAD also supports the current model for authority control by means of a controlled access point.

A similar approach is also evident in RDA: Resource Description and Access. RDA has to maintain compatibility with current cataloguing based on AACR2, but also look ahead to linked data. RDA therefore makes provision for “authorised access points”, equivalent to the current heading strings used in bibliographic records, but RDA also defines attributes by which entities can be identified and rich vocabularies for expressing the relationships between entities. As well as traditional “library” relationships, such as earlier/later names, or distinct bibliographic identities, RDA defines relationships more typical of archival practice, such as those between members of a family

or an individual and an institution. As well as facilitating machine matching, such relationships can support navigation and presentation of results.

As a consequence of these developments, many new fields are being proposed for inclusion in the MARC 21 Authorities Format. These will make it possible to explicitly record dates and events associated with a person, family or corporate body.

Explicit authority data can be exploited by machines to match authority records more efficiently. This process has been exemplified by the Virtual International Authority File (VIAF), a collaborative project initiated by Library of Congress, OCLC and Deutsche Nationalbibliothek. VIAF has successfully matched authority files from different national libraries and has also matched sample data provided by rights management agencies involved with ISNI.

A really exciting application of automated matching is when rich authority data can be reliably matched with bibliographic data. This is one of the potential applications of VIAF. It is also fundamental to the approach taken by the JISC funded Names Project. Names is led by MIMAS, at the University of Manchester in partnership with the British Library. The project is investigating the provision of web services for deduplication and disambiguation of names in the context of institutional repositories for Higher Education and grant making

bodies. Initial work has involved the deduplication of names associated with article level data (using Electronic Table of Contents). The project also aims to match these identifies with other authority files, including NACO. The project also has to develop interfaces and services which will enable authors to control their own identifies. The role of the professional cataloguer will move from data creation to data evaluation, assurance and service enhancement.

Libraries invented the concept of authority control. The semantic web will be built in part on accurate identification of persons, families and corporate bodies expressed as authoritative linked metadata. Libraries and rights management agencies are the biggest custodians of such data. In making this transition, much that is taken for granted will change. In a linked data environment far less significance attaches to a preferred form of name than in a linear catalogue. The scale of the task is such that it can only be accomplished through the concerted efforts of many partners. New partners may be engaged as the implications of linked data become clear and the benefits of control are more widely recognised, but this will also depend on new and efficient tools which can be integrated with existing workflows. This will not be easy but libraries do not have to do it alone.

Useful links

<EAC-CPF> Encoded Archival Context: Corporate Bodies, Persons, and Families
<http://eac.staatsbibliothek-berlin.de/>

FO:AF (Friend of a Friend)
<http://xmlns.com/fo-af/spec/>

Functional Requirements for Authority Data: a conceptual model
<http://www.ifla.org/publications/ifla-series-on-bibliographic-control-34>

International Standard Names Identifier (ISNI)
<http://www.isni.org/>

Interparty Project
<http://www.interparty.org/>

Communicating knowledge: how & why UK researchers publish & disseminate their findings
<http://www.jisc.ac.uk/publications/documents/communicatingknowledgereport.aspx>

Library annual report & accounts, 2008/09
<http://www.bl.uk/about/annual/2008to2009/index.html>

The knowledge network: British Library annual report & accounts, 2008/09

<http://www.bl.uk/about/annual/2008to2009/index.html>

Library of Congress Acquisitions and Bibliographic Access Directorate *Report of Fiscal Year 2008 (Fiscal Year Ended September 30, 2008)* <http://www.loc.gov/catdir/cpsol/aba08.pdf>

Names Project
<http://names.mimas.ac.uk/>

Repositories (statistics)
<http://roar.eprints.org/index.php>
<http://www.nostuff.org/ircount/index.php>

Statement of International Cataloguing Principles
<http://www.ifla.org/publications/statement-of-international-cataloguing-principles>

RDA: Resource Description & Access
<http://www.rda-jsc.org/>

United Kingdom Repositories: table of record totals
<http://www.nostuff.org/ircount/table.php?frequency=monthly&country=uk>

Virtual International Authority File: VIAF
<http://viaf.org/>

based on the presentation given at CIG authority control workshop 2009, London

Alan Danskin, Metadata and Bibliographic Standards Coordinator, British Library e: alan.danskin@bl.uk

The Online Computer Library Center (OCLC) first published the Dewey Decimal Classification (DDC) scheme online in 1993, with the launch of the current WebDewey interface in 2000. The site states that WebDewey is to be used to 'find a classification number appropriate for the item you are describing [and to] identify additional subject terms to use as access points.'. It offers value added features unavailable in the printed volume such as Library of Congress Subject Headings (LCSH) and is updated quarterly. A variety of pricing models are available which can make WebDewey a more economic option than the printed edition. The focus of this article will be the key features and drawbacks of using WebDewey for classification, and is written by a novice classifier.

Searching

The first problem encountered when using WebDewey is navigating the interface. The headings for the different services offered within OCLC's Connexion suite are not intuitive for a first time user. WebDewey is accessed via 'Dewey Services' along a set of navigation tabs and directs you to the main search page. The main search page offers a variety of search field options and separate search boxes for building search strings easily using Boolean operators. The search fields echo the printed version; although it is fairly obvious which field searches what, some practice is required to find the most appropriate search field for the particular topic or item.

Searching all fields unsurprisingly retrieves a large number of results. This approach is useful in

searching compound terms and for occasions where the actual class name might not be known (e.g. "commercial" rather than "business" law).

Searching the relative index using a simple keyword highlights some difficulties with the search, for example the term "dog" retrieves wider results than might be anticipated. Results include Management Measures (including note refers to "yellow-dog contracts") and Pest control (relative index terms include "Dog pounds").

The screenshot shows the OCLC Connexion WebDewey search interface. The search term 'dog' has been entered, and 19 records were found. The search results are displayed in a table with the following data:

Record Number	Classification Number	Subject Term
1.	331.894	Management measures
2.	363.78	Pest control
3.	599.77	*Canidae (Dog family)
4.	B 599.7721594	Dogs--vocalization, ...
5.	B 636.70831	Doghouses, ...
6.	B 636.708870284	Dogs--care and maintenance--pets--equipment, ...
7.	636.71	Breeds of dogs
8.	M 636.72-636.75	Specific breeds and groups of dogs

The top-level class for dogs as domestic animals (636.7) is not retrieved using this search term. Additional subdivisions and built numbers (indicated by the letter B) are retrieved in this search. The class number 636.7 is only retrieved using the keyword "dogs" (fifth result of 29) or using truncation dog* (where it is result 18 of 49).

When using single, broad keywords the search retrieves results you might not expect. These results are appropriate to the search field used and reflect the wide scope of DDC, however, they could initially confound a new user or novice cataloguer as it is not obvious (until you enter the record itself) why they have been retrieved. Using more focused search terms (as is recommended in the quick tips) returns more accurate results, but is rather limited in its scope and will often return no results when searching for more complicated concepts. This limitation is common to all schemes, regardless of format and can be solved through keyword reconfiguration. Greater familiarity with the scheme and its terminology would also reduce such erroneous retrieval.

Dewey number searching is a useful feature, but is limited as it searches only main class numbers and added built numbers. On the results page the display segments the number to show where division has occurred, which is useful for interpretation of built numbers and adapting the given built number.

Searching takes getting used to, especially regarding which fields to use

to retrieve relevant results, getting accustomed to what level of detail in a search is necessary and what will prove excessive. This problem may be due to the medium, as it is all too easy to forget that Dewey is a scheme capable of multiple subdivision, therefore when searching online it is imperative to employ the same keywording principles as you would when searching the printed volume, rather than assuming similar functionality to a generic search engine.

Browsing

Browsing offers a closer approximation to the searching process within the printed volumes. It proves very useful when using simple keywords or approximate phrases, for example it retrieves precise results for “dog” and “business law” which are easier to interpret than within the search interface. Browsing for “dog” shows “dog family” whereas the search facility gives the Latinate class name. Business law maps to its commercial law, but displays business law on the browse results page. These examples show how browsing might be easier for either an individual without specific subject knowledge or a novice classifier. Browsing the relative index is straightforward as “page up” and “page down” function offers navigation through the index and echoes the printed version.

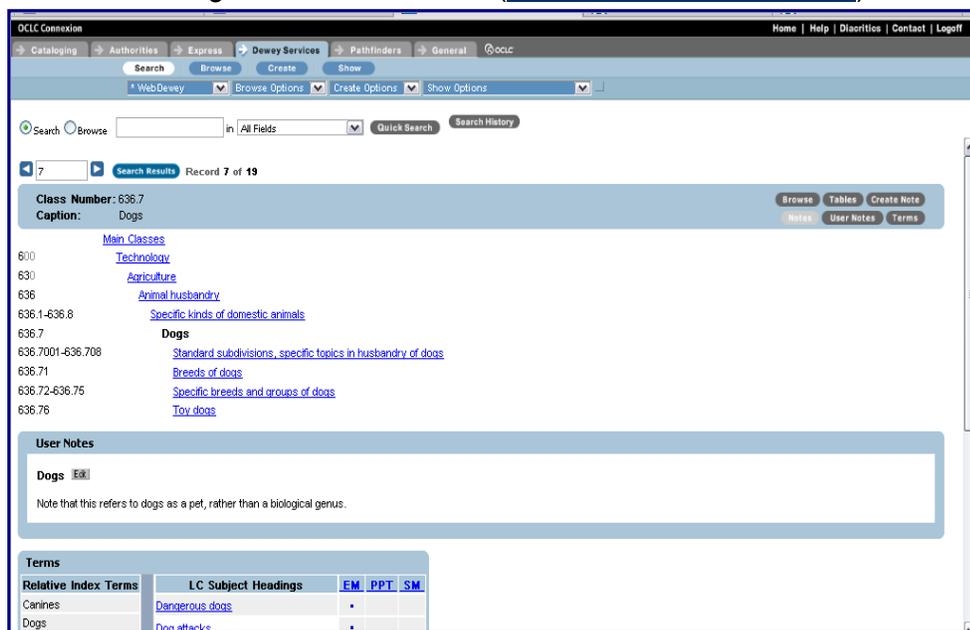
Results

Individual record pages are helpfully laid out, showing the term’s place in the hierarchy including those class numbers above and below it. These numbers are hyperlinked to enable browsing. It offers a good snapshot of the class number’s position within the overall hierarchy, but might not offer quite the same

wider perspective as the printed volume, especially for a novice classifier.

The use of boxes means that notes, relative index terms and subject headings are easy to distinguish. Added notation is clearly asterisked and the notes box provides a hyperlink to the relevant section, making the instructions easier to follow than the printed volumes. These hyperlinks open in new windows (or tab in Firefox), allowing simultaneous reference to the original number and the notation. This is a definite improvement on the paper version. Notes also include hyperlinks to other related or more appropriate class numbers. The individual record will also include any individual or institution notes which have been added. Such notes are useful for maintaining consistency in local cataloguing policy, they are easy to create and integrate in to the main sequence as a written note would be in the printed volume. It is possible to create institutional notes as well as notes for individual cataloguers.

LCSH appear on the record page to aid subject heading assignment, saving time and effort (it is worth noting that DDC-LCSH mapping was suspended for some time). However, these are only a selection of subject headings, so it probably does not match the DDC-LCSH mapping available through Classification Web (<http://classificationweb.net/>).



Navigation

The main navigation is not intuitive as it operates through a combination of tabs and drop down menus. As well as the format, the function of these navigation menus is not obvious, for example viewing a table is the “show tables” option within the drop down menu labeled “show options”. Helpfully opens in a new window to allow simultaneous main class and table viewing, but is not easy to find. Although not obvious, it is a very easy system to use once the user is accustomed to it and OCLC has created both quick and in-depth help documentation to support users. Navigation between search results records is straightforward, using the “view record jump bar”.

Workspace

The workspace is a temporary storage area to aid number building. Hyperlinks to tables and notation open in separate windows. The workspace sits at the top of the screen, in to which numbers from the main sequence and tables can be copied and pasted for building numbers. It makes keeping track of number building easier than in print.

Conclusions

The main stumbling block of WebDewey is that it is not an intuitive system to use, however, this is overcome through extensive help documentation and familiarity. The difficulties of searching (irrelevant results) are related to the problems of searching a complex classification system, rather than a fault of the interface as similar problems are encountered when using Classification Web. It is a very useful classification tool, it simply requires some patience whilst building up familiarity with it.

Further information on WebDewey:

<http://www.oclc.org/dewey/versions/webdewey/>

Elly O'Brien,

Information Resources Librarian ,
The Royal College of Surgeons of England.

CALL FOR PAPERS

The editor is accepting proposals for issue 159 of *Catalogue & Index*, which will be covering cataloguing, metadata and indexing within the film and image sector.

We are looking for:

- short vignettes (300-600 words)
- case studies of specific projects / roles (600-900 words)
- discursive essays on aspects of practice / experience (upwards of 900 words)

Suggested topics could include, but are not limited to:

Ideally, we would like to represent all the major sectors of the LIS community: academic, commercial, public, and special. We would also be happy to hear from students, lecturers and para-professionals.

This issue will be published electronically and be made available through the CILIP Cataloguing and indexing Group website

<http://www.cilip.org.uk/get-involved/special-interest-groups/cataloguing-indexing/Pages/default.aspx>

C&I is used as a promotional tool at future events, so if accepted, your work will reach not only the existing readership of *C&I*, but a wide range of others interested in cataloguing, indexing and metadata.

In the first instance, please send your name, contact details and proposal to the editor, if accepted the final submission date for articles will be confirmed via email.

Email the editor: penny.robertson@sqa.org.uk

I attended this workshop organised by the Catalogue and Index Group of CILIP on 23rd October at Ridgmount Street. I was particularly keen to attend as we are currently doing a lot of retrospective AC work at Leeds Met. I'm also now involved with teaching the Organisation and Retrieval of Information course to the Masters in Information Studies students at the University, and I felt I might need a refresher!

The workshop consisted of four presentations with time for questions and discussion. It was well attended, with delegates mainly from various academic libraries, but some from public and special libraries, which perhaps reflects where most "traditional" cataloguing is now done.

The first speaker was Anne Welsh of University College London. Her presentation was titled "Resource discovery: authority control from the user perspective". The talk was bright and breezy, very much a personal view, and took the concept of authority control outside the library catalogue to the Internet, Amazon and Google, and institutional repositories. It was a good start to the workshop, reminding everyone that this apparently abstruse and even arcane discipline in fact has wide-reaching application.

Helen Williams of the London School of Economics then spoke about "Retrospective Authority Control". This was very much a report of the project recently completed at LSE to tidy up their authority files. I found this particularly interesting as I am trying to do much the same thing at Leeds Met, albeit on a very ad

hoc basis! Helen concluded that the project was very worthwhile, though extremely hard work in spite of using an external contractor.

Hugh Taylor from the University of Cambridge spoke on the Library of Congress/NACO Authority File. This was an entertaining and enlightening session, with a welcome emphasis on the end user, not just the technical services staff.

The last speaker was Alan Danskin of the British Library, on "Automation and the Future". Alan highlighted the very large gap between the number of published resources per year and the number of authority records produced, and the widening numbers and types of institutions involved in controlling information. He asked whether we need to rethink the process, and whether Authority Control is in fact being applied at the wrong point, i.e. at the end of the process instead of the beginning. He touched on developments in MARC, on RDA and the concept of the International Standard Name Identifier, and asked how these would further affect our current ideas of AC. This session probably raised more questions than it answered, but was fascinating.

I enjoyed the sessions very much and have found them useful, especially as an update in where AC seems to be going, and also giving me information which helps to emphasise the importance of Authority Control in a large collection – possible ammunition for the future. There was also some information which I shall be

passing on to our university's Repository Development Officer. I felt it was aimed at the experienced librarian rather than trainees and students, so fulfilling a need not met by the CILIP training courses in cataloguing. My only criticism was the description of the afternoon as a workshop, when it was actually a series of presentations – but none the worse for that! I would certainly be very happy to attend similar events on specific Cataloguing subjects, and I hope some will take place in the North of the country.

Kate Bunting, Senior Bibliographic Services Librarian, Leeds Metropolitan University.
e: k.bunting@leedsmet.ac.uk

Book reviews

“E-Journal invasion: a cataloguer’s guide to survival” by Helen Heinrich. Binghamton N.Y.: Haworth, 2007.

This is a useful guide if you want to know a brief history of electronic journal cataloguing, coupled with a case-study as experienced by the author. Serials cataloguing is a fairly rarefied field of librarianship; electronic journals cataloguing, as a subset, even more so. There are more ways to catalogue a serial than ways to have hot dinners, which is why there is such a variation in catalogue records for the same title – if a definition of a serial can be agreed upon in the first place.

Heinrich steers through these muddy waters with ease, offering a summary of the cataloguing basics for serials. She covers the MARC21, MARCXML, MODS, Dublin Core and ONIX schema. AACR2 rules are also summarised, particularly chapters 9 and 12, which are concerned with electronic resources and continuing resources respectively. It is noted that AACR2’s replacement, RDA, may offer a different cataloguing solution. The pros and cons between the single record and multiple record approaches are discussed, along with the CONSER standard. The cataloguing aspect is covered with reference to clear and succinct diagrams. This takes out the headache of having to reconcile AACR2 with MARC21 and any local practices, for example, with illustrations from different websites

to show different schema. There is even a useful table to describe the variations (and similarities) between a serial and an integrating resource – basic stuff, but useful to have lain out in one place.

One major issue covered is that of the ever increasing numbers of electronic serials that are available, e.g., through aggregators and bundled packages. It is made clear that the traditional method of cataloguing any resource is simply not able to keep up with the volume of journals that are now available. This is where e-journal management systems are described, such as TDNet and Serials Solutions. These commercial products are taken in an impartial manner, which allows the reader to draw their own conclusions as to their worth.

The author’s experience of implementing Serials Solutions’ MARC record service in her library catalogue is given as a case-study. This is detailed, starting with the preparation and testing, working through the problems encountered, detailing the actual implementation, subsequent clean-up file and continuing updates. Questions raised (and answered) include easily-overlooked details in the general plan, such as the amount of time needed for the preparation and ongoing maintenance. Another case-study would have been useful, allowing for a different perspective.

The book finishes with a summary of the issues faced – and will face

in the future – and some of the other opportunities that the library world needs to embrace so as not to be superseded by the likes of Google Scholar as the research tool of choice. The shift in emphasis in cataloguing from description to accessibility is made clear, alongside the need for comprehensive metadata to link through from any search to the resource.

This is a useful summary and guide to have on the serial cataloguer’s bookcase. Despite the simple language used, the issues covered are complex. The author has a firm grasp on the challenges faced when cataloguing electronic journals, which can aid not only the cataloguer but also the acquisitions and the reference librarian.

Natasha Aburrow-Jones,
SUNCAT Project Officer, EDINA
Service Delivery - Bibliographic and
Multimedia Services.

Usage statistics of e-serials, David C. Fowler (ed.). Haworth Press, 2007. Simultaneously co-published as *The Serials Librarian*, 53 (Supplement 9). 297p. 978-0-7890-2988-1

'Lies, damned lies and usage statistics – what's a librarian to do?' asks the title of one article in this collection, and it is no doubt a question that many a librarian has found themselves asking when trying to make sense of an overwhelming mass of usage data.

This volume attempts to provide some answers in the form of seventeen journal articles which cover a mix of case studies of how individual libraries manage usage statistics and more theoretical considerations of the potential and limitations of usage data.

The case studies cover many of the practicalities of implementing a usage statistics programme within a library, including the cost and time involved in collecting and processing usage data, and the thorny issues of data reliability and comparability. They also discuss the measures which can be produced from usage data, such as calculating cost-per-use for different resources, and how they can be used to support collection management. Whilst most concentrate on the simple reports based on vendor-supplied statistics, there is also some coverage of more sophisticated techniques such as deep-log analysis of data from local library

systems.

The more theoretical articles consider the advantages and disadvantages of relying on usage reports, in particular looking at the extent to which usage can (or cannot) be related to value. There is discussion of the potential of combining quantitative usage data with more qualitative assessments to gain better insight into user needs and behaviours, alongside warnings of how over-reliance on measurement could distort both libraries' collection management decisions and the scholarly communication supply chain itself.

Although this volume aims to cover a broad range of issues, the perspective is largely that of US academic libraries, and the brevity of the articles means that some complex issues do not get the depth of treatment they deserve. There is considerable repetition between several of the case studies and as some of the articles are now a few years old much of the detail is out-of-date, although the broader issues remain pertinent.

But despite these limitations, this volume nevertheless provides a useful introduction to many of the key issues in collecting, analysing and interpreting usage statistics for e-resources. Practitioners new to this area should find the case studies useful, and for managers seeking an overview of how usage statistics can support collection management decision-making, it

provides a useful summary of both the potential and limitations of usage statistics.

Anna Grigson, Assistant Digital Resources Librarian, University of Westminster



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 13.50 for non-members.

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>