

Catalogue and Index

Periodical of the Cataloguing and Indexing Group,
a Special Interest Group of CILIP, the Library and Information Association



Editorial

March 2017, Issue 186

Welcome to C&I issue 186. My name is Deborah Lee, and I have recently taken on the role of co-editor of C&I, to work alongside Dr Karen Pierce. I am the senior cataloguer at the Courtauld Institute of Art, and have also recently submitted my PhD thesis about classification at City, University of London. I always love reading C&I so I am really looking forward to my role as C&I co-editor.

Issue 186 has two interrelated themes. The first theme continues our coverage of the 2016 CILIP CIG conference, with the inclusion of more conference papers. Amy Staniforth begins by outlining her work designing and delivering "beginners' cataloguing training", and argues why such training is sorely needed within the library and information sector. Suzanne Barbalet and Nathan Cunningham write about using classification for new and novel forms of data, and outline their pilot study which utilises UDC at the UK Data Archive. Ahava Cohen discusses multilingual, multiscript authority control, and the project to set up a multilingual and multiscript authority database in Israel. Anne Welsh discusses FRBR, BIBFRAME and relationships in cataloguing, with particular focus on the cataloguing of items as objects.

The second theme follows on from the topics of the conference: using current projects and working practices to discuss the future of cataloguing. Elly Cope describes a project which uses Google Analytics to analyse how users interact with the University of Leeds Library's discovery layer. Diana Palmer describes a project at the Heinz Archive and Library, National Portrait Gallery, which focuses on cataloguing books about photographers. Finally, Alison Felstead writes about her career in cataloguing, taking the reader on a journey through the cataloguing past, her work as Head of Resource Description at the Bodleian Library, then through to her current role. Alison links the cataloguing past to the cataloguing future, sharing her thoughts about the future of cataloguing and the cataloguing profession. I hope you enjoy this issue of C&I.

Karen F. Pierce: PierceKF@Cardiff.ac.uk
Deborah Lee: Deborah.Lee@courtauld.ac.uk



Contents

2-9 We need to talk about cataloguing, by Amy Staniforth

10-14 New tricks? Old tools to solve new problems, by Suzanne Barbalet & Nathan Cunningham

15-19 Each of us has a name: creating a multilingual, multiscript national authority database, by Ahava Cohen

20-29 From WEMI to WI to WII: FRBR, BIBFRAME & the 21st century bibliographic model, by Anne Welsh

30-32 Google analytics for discovery layer insight, by Elly Cope

33-34 Books on photographers: cataloguing at the Heinz Archive & Library, by Diana Palmer

35-38 Looking back, looking forward: reflections from a career in cataloguing, by Alison Felstead

39 Book review, by Anne Welsh

I don't need to convince anyone reading Catalogue & Index that good metadata is at the heart of a good library service but I expect that we have all heard remarks from colleagues that suggest this isn't universally understood. "Isn't it all online now?" "it's just the ISBN that we need" and a personal favourite "can't we just put it in a note field?"

In this paper I argue that we need to talk about cataloguing with non-cataloguing colleagues and I describe an unusual cataloguing 'conversation' I have been having with a colleague that earlier this year, turned into a beginner's cataloguing session for librarians. I will conclude with our brief assessment of the exercise and hopes to develop further conversations in the future.

Why a conversation?

Yes, there are cataloguers who focus time we no longer have on detail while the big picture threatens to sideline our expertise but ultimately, wherever metadata comes from and however it is presented, it is identified and created, refined, accumulated, and shared by cataloguers to lay the foundations of library services. Without metadata you have a warehouse of books at best. To create a resource that people can search, we all – our collection managers, customer services, systems technicians, academic engagers, and information literacy tutors – need bibliographic metadata.

Not all records are created equal...





Important as it is, however, I must admit that our metadata is of varying quality.

As well as all the unknown unknowns (I learned on the job so there are many of these) I do know that we have some fullish records that effectively describe and identify aspects of works, expressions, manifestations and occasionally items, and some that clearly do not.

Working on -Llon Gyfarchion. (995825623402418) , Created by import (09/07

LDR			00290nam#a2200121la#4500
001			995825623402418
005			20090821164724.0
008			090817s9999####x#####000#0#und#d
035			\$\$a (WIAbUW)582562-aberdb
245	0	0	\$\$a Llon Gyfarchion.
260			\$\$a Dinbych : \$\$b Wasg Gee , \$\$c 1977.
300			\$\$c 145x175mm.
852			\$\$b holibrbr1 \$\$h Z42.L7
945			\$\$a rbrdata

Two rare book bibliographic records from our Library Management System (LMS) Alma (1)

005			20160531123749.0
008			160527s1876####enka####r#####000#0#eng#d
040			\$\$a WIAbUW \$\$b eng \$\$ rda
041			\$\$a eng
050	0	0	\$\$a E168 \$\$b M2
	100	1	\$\$a Manning, Samuel, \$\$d 1822-1881 \$\$e author
	245	1 0	\$\$a American pictures drawn with pen and pencil. \$\$c By the Rev. Samuel Manning, LL.D., author of "Swiss pictures
	264	1	\$\$a [London] : \$\$b The Religious Tract Society, 56 Pasternoster Row, 65 St. Paul's Churchyard, and 164 Piccadilly,
	300		\$\$a 224 pages : \$\$b illustrations (black and white) ; \$\$c 29 cm
	336		\$\$a text \$\$2 rdacontent
	336		\$\$a still image \$\$2 rdacontent
	337		\$\$a unmediated \$\$2 rdamedia
	338		\$\$a volume \$\$2 rdacarrier
	505	0	\$\$a How it strikes a stranger -- St. Louis to Denver -- Amongst the Mormons -- The new El Dorado -- The Yosemite a -- The Empire State -- Philadelphia and Washington.
	561		\$\$a Copy 1: Bookplate of George Ernest John Powell, Nant-Eos (Aberystwith) \$\$5 WIAbUW
	561		\$\$a Copy 1: Bookplate of University College of Wales, Aberystwyth. General Library. \$\$5 WIAbUW
	561		\$\$a Copy 1: Inscription on title page G. E. J. Powell \$\$5 WIAbUW
	563		\$\$a Publisher's pictorial binding stamped on spine and upper cover with gold and black. The U.S. Capitol building, G train on a trestle bridge are featured in gold on the cover. All edges gilt.
	650	0	\$\$a Travel
	650	0	\$\$a Mormons
	651	0	\$\$a United States \$\$x Description and travel

Two rare book bibliographic records from our Library Management System (LMS) Alma (2)

I'm sure that we are not unique and that everyone has their shameful, or "non-priority" records, but it is important that colleagues who have to make priorities and decisions know about these records and the potential consequences of quality or lack thereof on borrowing statistics, especially when space is such a pressured commodity. And it isn't only our legacy metadata that our colleagues need to know about. While a lot of our work is not "original" cataloguing, the now numerous methods of obtaining metadata, and often altering it, make untangling problems seemingly remote from the metadata itself, increasingly complex.

We've had issues where \$z in the 020 has been preventing academics linking to thousands of subscribed to e-book records in our reading list system, for example, and where librarians have been surprised by partial search results for special collections because names of donors have been idiosyncratically added to different bibliographic fields.

Even these few examples demonstrate that when we do have to create metadata or diagnose and fix metadata issues we need skilled cataloguing librarians who know what they are looking at or where to go to find out. And there are less and less of us as services have increasingly been outsourced and automated.

Institution	No. Cataloguing Staff
Aberystwyth	2 x P/T
Bangor	1 F/T 1 P/T + occasional assistance from frontline staff
Cardiff Met	0.8 + occasional assistance from frontline staff
Cardiff Uni	5 FT 2 P/T 2 F/T <u>Acq Asst</u>
Glyndwr	<u>Univ Lib</u> , Deputy Lib & 3 Learning Resource Advisors catalogue stock for their subject areas.
NLW	35 F/T 10 <u>Asst</u>
RWCMD	2 x P/T
South Wales	1 F/T
Swansea	1 F/T 2 <u>Acq Asst</u>
Trinity St David	2 x P/T

Extract from Jane Daniel's cataloguers survey of Wales Higher Education Libraries Forum (WHELF) institutions

So, if there is one cataloguer at an institution and often little awareness of metadata issues, then we are in an uncomfortable position (for cataloguers, perhaps?) in that we are left with the job of promoting what we do.

One case in point for us in Aberystwyth is authority headings / authorised access points. Having watched colleagues' faces glaze over when I talk about authorities I am then infuriated if someone asks me to magically bring together all the records for resources on a certain subject, or expects me to recommend paying for third party services to improve our discovery. Going back to square one I recorded the results positions of some recently added books before and after removing authorised access points, contents and summaries, giving me clear evidence that good metadata is a service to our users and not a quirky cataloguing tic.

Hatherley, Owen *The ministry of nostalgia* London : Verso, 2016

Search term	Results position in Primo 02.11.16 (with full record)	Results position in Primo minus the 505 & 520 fields (contents & summary)	Results position in Primo minus 505 & 520 & authorised access points
Hatherley	1 of 4	1 of 4	2 of 4
Nostalgia	11 of 419	30 of 419	60 of 419
Nostalgia, 20 th century	6 of 110	9 of 110	0 of 109
"Lash Out and Cover Up" Ch.1 title	1 of 9	0 of 8	0 of 8
"Keep calm and carry on" phrase in summary	2 of 5	0 of 4	0 of 4

Extract of business case for metadata: authorised access points / headings

So there are important business cases to be made. But – and it is a big but – I would also argue that we have a responsibility to talk about cataloguing...because it is fun.

Maverick Metadata (or, Titans of Tat, Captains of Crud...)

Not fun for everyone, granted!

At Aberystwyth we have 1 FTE cataloguer and our time is spent almost entirely on bread and butter new print and electronic books cataloguing. We don't have a special collections librarian and as well as a nagging guilt about the non and poorly catalogued rare books in our care I have become increasingly aware of how difficult it is to switch focus from mass copy cataloguing to rare books.

My colleague Simon French came from book selling to libraries and was a graduate trainee when he first worked with us on cataloguing. He wasn't impatient with it or me and, in fact, appeared to like rather than loathe the slow detailed attention needed, and clearly loved books and describing them. Luckily for me Simon was recently appointed a subject librarian and was amenable to my suggestion that we try cataloguing...for fun. We decided on a weekly lunchtime cataloguing session that allows us to choose uncatalogued items we like and gain special collection cataloguing skills with the happy by-product of adding material to stock. Whilst serious about books and making them accessible our silly names (see above heading) reflect our approach and the fact that we know and embrace the 'nerdiness' of our enterprise – friends remind us of this frequently.

Although it is a problem that I do not have time for day-to-day rare book cataloguing, Maverick Metadata (MM) also reflects the power of doing something on our own time. We don't necessarily pick the prettiest, shiniest, most-relevant-to-our-current-syllabi, books and even though our pace is meandering at best, we take a perhaps perverse pride in endowing an unloved book with metadata and making it discoverable.

Along the way we do enjoy some detective work, as I know cataloguers everywhere do, when they have the time. We were chuffed when this bookplate: \$a Carolii ac Mariae Lacaitae Filiorumpque, Selham, Sussex, eventually led us to a Library of Congress heading: Lacaita, C. C. \$q (Charles Michael), \$d 1853-1933 \$e former owner. I haven't described Bibles before and on another occasion cataloguing a beautifully bound Moses Mendelssohn collection we were looking for a heading to reflect his work on the Five Books of Moses and were delighted to come across this after stumbling around in Library of Congress Subject Headings (LCSH): \$a Bible. \$p Pentateuch. \$l German. \$s Mendelssohn.



We do also get out, sometimes! On the Information Services away-day coastal path walk in July we had Detectorist¹ style photos taken for my CIG conference report on the workshop – and brought along our own (different edition) copies of Carter's ABC for Book Collectors.

Pleased with our own progress, I was fired up by CILIP 2015 in Liverpool (thanks to a CIG bursary!) but also curious about how little interest there was in metadata sessions. They were the least attended of the sessions I went to and it struck me that cataloguing and metadata appear not to be FOR all librarians and this, together with the CILIP Wales 2016 conference call for papers and workshops made me persuade Simon that we should consider sharing our experience.

Beginners' cataloguing...

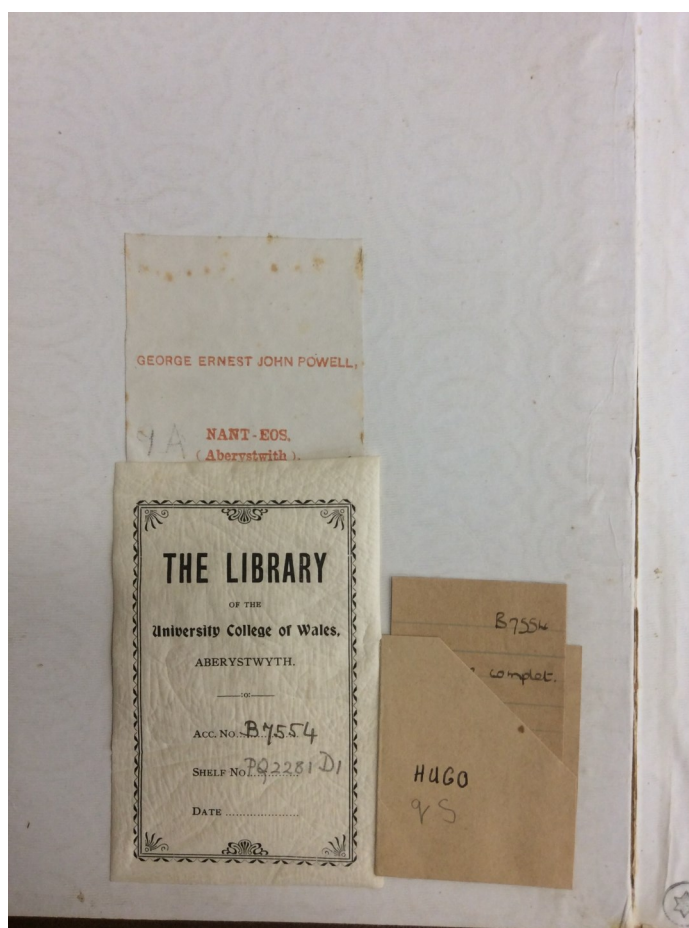
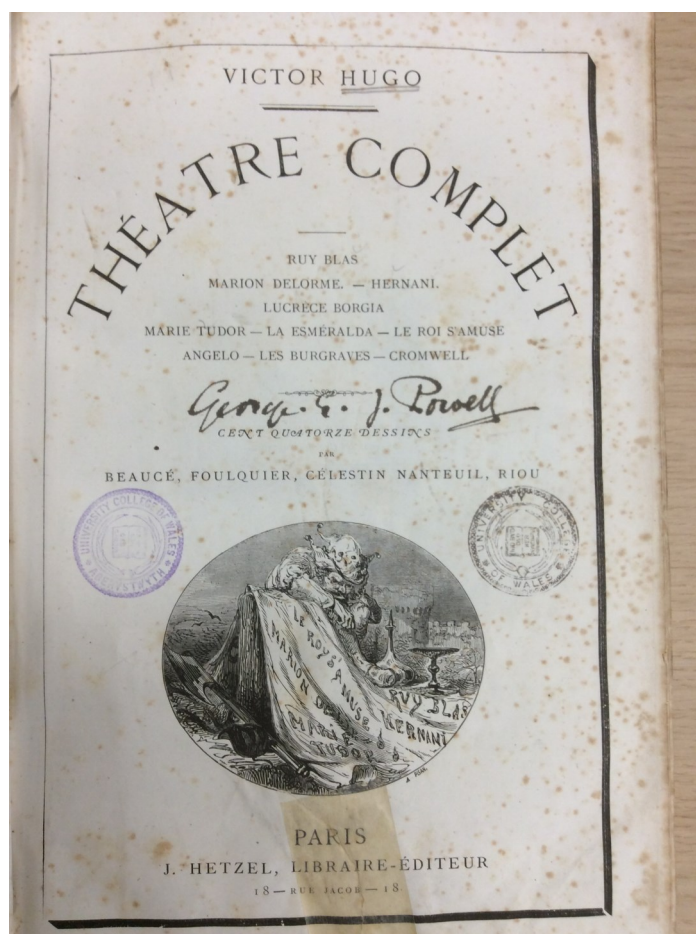
So, we designed a session:

- 1 hour long session - to fit into CILIP Wales & our department's training slots
- Aim: to create bibliographic, holding & item records for a rare book live in Alma, the new WHELF LMS
- Resources used: the book, hand outs of title pages & book plates, basic templates, and the web resources LCSH, Worldcat, Classweb, COPAC, Wikipedia, & Marc 21
- Bib template: described as the tip of the iceberg and illustrated this by looking at the Marc 21 505 formatted contents note

¹ Detectorists – a BBC comedy about two metal detectorists, well worth a watch: <http://www.bbc.co.uk/programmes/b06l51nr>

- Context: used the idea of Work, Expression, Manifestation, and Item to identify different bits of bibliographic data on the item and title page hand-outs (title, language, pagination and inscriptions, for example)
- Item: briefly introduced and contextualised, then passed around (with handling guidelines) and record was created together (as much as possible) in a pre-constructed template in Alma on a big screen.

For an hour long session it took a lot of preparation. We had to find the right book – it had to be interesting but not too interesting because of time constraints – and then practice the session to squeeze it into an hour with time for questions and the unexpected.



An example of one of the items catalogued – each had bookplates and inscriptions and stamps

We gave the workshop three times, twice in Aberystwyth and once at the CILIP Wales conference in Swansea. The audiences came from across different library areas (and a bit beyond) with subject librarians, an archivist, managers, students, collection management and colleagues from customer services, acquisitions, and systems all choosing to come along.

Beginners' cataloguing ...some feedback (from Aberystwyth participants):

I think an interesting item is really important...

maybe DIS [Department of Information Studies] students could be involved at some point? There may also be colleagues from academic departments who are interested.

I think it might also be good to expand a bit more on the resources you use when researching dates, etc, as people might not be familiar with WorldCat, etc (I know I wasn't).

Thanks - like I said, this is really interesting. I love the "abstract" idea of the resource and the content, then boiling it down to the actual physical book.

Thanks very much, it was really interesting! Maybe I'm more geeky than I thought.

I found the "common" headings part of the system really interesting, choosing search terms which will bring back the best possible results.

I was surprised how manual the system is for particular items, and how much everything has to link to other catalogues.

The message I took away was "Aberystwyth has some special books" "there's a lot more to cataloguing than people think", and "Alma is not something you can pick up and run with, neither is cataloguing".

Beginners' cataloguing...our reflections:

We felt that although an hour was clearly not enough colleagues appreciated the complexity of cataloguing and demonstrated an interest in the material (item-in-hand) and abstract/philosophical (shared vocabularies etc.) nature of cataloguing. It also ended up being a good illustration of the tension between template work and the constant decision making involved in cataloguing which is difficult to explain to managers.

Given people's interest once working on an item it struck me as odd that for many colleagues as well as students, metadata seems to just 'exist out there' without creators or context. Ultimately, what the session showed us was that cataloguing isn't part of ongoing conversations, library or otherwise, unless we put it there.

Next:

Maverick Metadata has gone back to its roots after the heady glamour of touring to Swansea for CILIP Wales 2016 and we continue to come across dusty gems and think through all sorts of ways of describing items and identifying key metadata to share with users and colleagues.

We would consider a session beyond Information Services as we do have Archives Administration and Librarianship postgraduates in Aberystwyth and quite a few other libraries too, but we would also like to propose a smaller multiple session course where each participant catalogues a book of their own.

I am scheduled to give a departmental talk on authorised access points/headings and classification too, as it is becoming increasingly clear to me that these two areas of metadata application are precisely where my work overlaps with that of colleagues and I need their experience and interest to tackle legacy issues and prepare for sharing our metadata much more widely in the future.

Given work commitments I may manage little of the above but our Maverick Metadata conversation convinces me that enthusiasm is not something to bury under professionalism; we need both if what cataloguers do is to remain at the heart of our libraries in the future.



An important characteristic of classification schemes is that they are stable structures so it might seem something of an anomaly to consider the introduction of a classification scheme into a data archive that is preparing to meet the challenges of new and novel forms of data (NNfD), particularly when such a system of collection organisation has not been used before. However, the very strength of a classification scheme is to “act as a model and a map of the domain” (Broughton, 2016: 339) and in this role it appears to be a useful tool to assist with organising and browsing new subject domains that NNfD may introduce us to and to complement keyword control of content of collection materials by thesauri. While generally favoured as a ‘mark and park’ tool classification schemes have been used online to organise topics for gateways such as Intute¹ or for the organisation of self-deposit documents in the Research Repository schemes² so it is not a path untrodden.

‘Future-proofing’ Access to NNfD

At the UK Data Service we prefer to refer to new and novel forms of data (NNfD) and ‘smart data’ rather than ‘big data’. Smart data is wide data (high variety), not necessarily deep data (high volume) and it is made up of “feature-rich content and context (time, location, associations, links, interdependencies, etc) that enable intelligent and even autonomous data-driven processes, discoveries, decisions, and applications (Borne, 2016).



Fig. 1: Smart Data

¹ <https://digital.humanities.ox.ac.uk/project/intute>

² For example <http://repository.essex.ac.uk/>

While ‘big data’ has been used by commercial organisations for market research for many years the use of such data for social sciences and humanities research is relatively new. As we acquire NNfD our role as a data service is becoming more proactive. The UK Data Service and the South African data service DataFirst have jointly undertaken a research project³ that provides a model for social science research. This research on fuel poverty analyses government data in two countries and is a particularly good example of data services exploring the potential of smart data. In turn this work provides some insight into the metadata that best supports data discovery systems in the future.

Our staff have been encouraged to acquire additional skills. They now have the skills to enable them to identify and scale the data for access, linking and analysis and to ensure that data conforms to Research Data Management (RDM) principles which underpin all UK Data Service archiving and curation work.

We are aware that our traditional clientele of academic researchers, policy makers, educators and students will broaden in the future and that their research needs may not be limited to what can be discovered in an online catalogue. It will be important to create user profiles to explore what aspects of NNfD users from outside our traditional community may require and consider the best way we might meet their future needs. For example, the Natural Environment Research Council (NERC) engages a similar multidisciplinary research network, makes accessible cutting edge quantitative research and its vocabulary service needs to be flexible enough to respond to new policy issues. Here topic access has proved useful as an initial point of entry for a cross-disciplinary search.

The screenshot shows a search interface with a blue header bar containing the word "Search" and buttons for "Search", "Clear", and a help icon (?). Below the header, the interface is organized into several sections:

- Free search:** A text input field with a help icon (?).
- Disciplines - Parameter groups:** A dropdown menu showing "All", "Administration and dimensions", "> Administration and dimensions", "Atmosphere", and "> Atmospheric chemistry".
- Discovery parameters:** A dropdown menu showing "All", "Acoustic backscatter in the water column", "Acoustic noise in the water column", "Active seismic refraction", and "Air pressure".
- Cruise/Station name:** A text input field with a help icon (?).
- Projectname:** A text input field with a help icon (?).
- Datasetname:** A text input field with a help icon (?).
- Sea regions:** A dropdown menu showing "All", "World", "> Arctic Ocean", and "> > Baffin Bay".
- Waterdepth (m) from:** Two text input fields for "from" and "to" with a help icon (?).
- Originator:** A dropdown menu showing "All" with a help icon (?).
- CDI partner:** A dropdown menu showing "All" with a help icon (?).
- Country:** A dropdown menu showing "All" with a help icon (?).
- Access restriction:** A dropdown menu showing "All", "academic", "by negotiation", and "commercial charge" with a help icon (?).
- Instrument type:** A dropdown menu showing "All", ">2000 Hz top-bandwidth single-channel seism", ">2000 Hz top-bandwidth sub-bottom penetrato", and "1000 Hz top-bandwidth multi-channel seismic" with a help icon (?).
- Instrument depth (m) from:** Two text input fields for "from" and "to" with a help icon (?).
- Platform type:** A dropdown menu showing "All", "aeroplane", "autonomous surface water vehicle", and "autonomous underwater vehicle" with a help icon (?).
- Measuring area type:** A dropdown menu showing "All" with a help icon (?).
- Temporal resolution:** A dropdown menu showing "All" with a help icon (?).
- Date (yyyymmdd) from:** Two text input fields for "from" and "to" with a help icon (?).
- Duration:** Two text input fields for "from" and "to", followed by a "Unit" dropdown menu showing "Hours" with a help icon (?).

Fig. 2: Example of a Topic Search

³ <https://ukdataservicesmartenergydata.wordpress.com/> Accessed 17/02/2017

It was the need for such flexibility that led us to consider a proposal to ‘future proof’ our subject category organisation by introducing a classification scheme to reduce legacy work involved in updating a flat list of subject categories or topics.

Why has subject or topic access been an important discovery tool for the UK Data Service? Data archived for the purpose of secondary data analysis have a particular set of discoverability issues that is independent of changes in the range and variety of data available for social science and humanities research. For example to ensure the validity and reliability of the survey methodology the concept attached to survey questions and scales in social science research in particular will not always be apparent. A high proportion of survey questionnaires embed standard scales to develop a measurement of a variable in the study. Scales may not be described in the documentation nor easily identified in the questionnaire. Even if identified the variables themselves may not appear relevant to the ‘study’ topic until the research design is described in the resulting publications. The range of indicators of ethnicity is one example. The UK Office for National Statistics recommends ‘country of birth’, ‘nationality’, ‘language spoken at home’, ‘skin colour’, ‘national/geographical origin’, and ‘religion’ as indicators of ethnicity. Thus we index our studies⁴ to variable level applying keywords from our thesaurus Humanities and Social Science Electronic Thesaurus (HASSET) and supplement subject access by also allocating subject category terms from a separate controlled vocabulary.

An acknowledgement of these discoverability issues can be found in the Data Documentation Initiative (DDI), the international standard tailored to describe data. DDI provides both a keyword field, which we populate with HASSET keywords, and an optional subject field.

Our subject controlled vocabulary is a flat list which requires periodic review to ensure that topics reflect new areas of funded research and current policy issues. If we managed it by using a classification scheme to introduce structure into this controlled vocabulary then editing the list would be a simple task. A pilot study was undertaken in 2014 - 2016 to assess the feasibility of classifying the whole data collection before the size of collection made this task uneconomic.

Research data in the digital age is growing exponentially (Corti et al 2014:1). It appears that the application of exploratory analysis using NNfD will become an integral part of the process of curating data for social science research and will impact on the way we manage our resources and on the services we provide. The UK Data Service has developed expertise in automatic indexing and improved the quality of our thesauri. The UK Data Service platform that is nearing completion will provide data analysis features and harness existing knowledge organisation resources to create a vocabulary service. It will not be an exclusively open data platform. The ‘data lake’ will be a secure repository capable of storing raw data in any format and will facilitate variable searches to the lowest possible level of granularity. Support will be provided for users with expertise ranging from the student to researchers who want to develop their own bespoke data tools. All will require efficient and reliable access tools, not only to locate variables but also sources of open raw data.

Given that a simple keyword search box is the way most searches are undertaken, and with increasingly large recall, attention has been drawn to the problem of subject searches. The question is posed “are we too obsessed with the notion of providing access to everything at the expense of the quality of the results?” (Tay, 2016: 113). Can we enrich such searches with tailored metadata?

⁴ We allocate a study number to each set of data and its documentation as it is processed. It may be either quantitative, qualitative or historical research. It is a consecutive number that retrieves all datasets and documentation relating to the particular research project.

The Pilot Study

We chose Universal Decimal Classification (UDC) for the pilot study as it is the scheme of choice in the UK for large special media libraries, but most importantly because it is an analytico-synthetic classification, that is its constituent parts can be analysed or parsed in the process of information retrieval. Where a specific new topic has not been included in the main tables it can be covered with the addition of an appropriate extension from the auxiliary tables. In addition, the notation allows the linking together of numbers from the main class. We arranged 'studies' for processing by the main subject category allocated to each study in our current system. Series and virtual 'studies' were excluded from this pilot study sample, as they are easier to classify, so we selected non-series 'studies' for classification in order to gain the best estimate of the cost and time the task would take. Thus, in the main, efficiencies were made in the range of classification numbers required.

It was decided to make the codes as specific as possible until we knew how many auxiliary numbers would be required. We found the option to build numbers to include extra subjects particularly useful. When required we can retrieve these secondary subjects and allocate the data set accordingly to the extra or new category while still retaining all the required information in the full classification number attached to the study.

We included 'form' auxiliaries to explore the use of the notation to link data to resources that may be useful for training and outreach purposes; for example case studies, illustrations/photos or bibliographical references. We also added auxiliaries of 'place' and 'time'. The latter, we thought, would be useful for easy identification of our historical data.

We created guides for the most popular UDC numbers and found that by the end of fourteen weeks a part-time employee, working 15 hours a week, completed the classification of 4,289 'studies' Now, with access to UDC Online, the results are even more promising. Reviewing the content of one existing subject category via the code label allows for the identification of significant sub-categories which could form a new category in their own right. Within the subject category of 'Society and Culture' below we can see that 'Leisure, Tourism and Sport' is emerging as a significant new category.

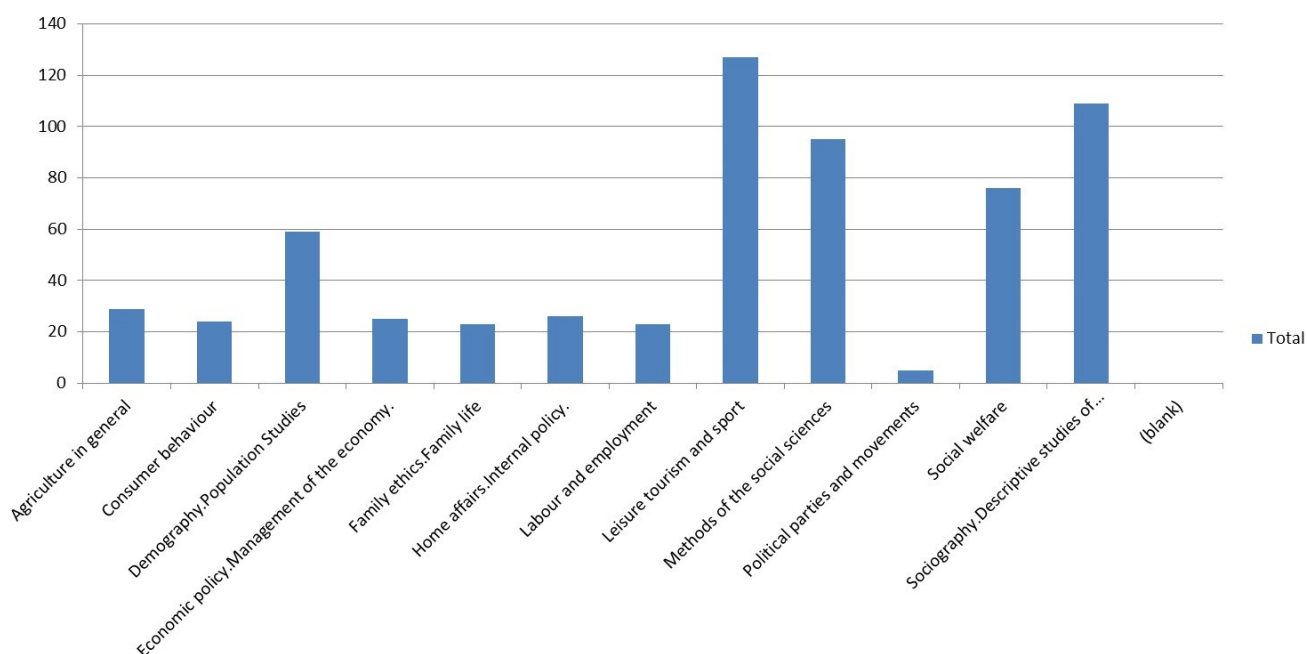


Fig. 3: UDC Notations Distribution for the Subject Category 'Society and Culture'

'Big Data' and 'Big Metadata'

While there is a popular view that 'big data' implies 'big metadata' currently there is no evidence to suggest that NNfD will fundamentally affect social science and humanities research methodology and that, in turn, the need for thesauri and controlled vocabularies will be replaced by Google style algorithms. The example of the Google Flu Trends research reflects the initial excitement.



Fig. 4: Flu Trends (Source The Guardian 27 March 2014)⁵

Researchers warn of 'big data hubris' (Lazer, 2014). Google's own autosuggest feature may have driven more people to make flu-related searches and misled its Flu Trends forecasting system.

The future is in the making and we have yet to see exactly what discoverability issues will arise as researchers explore new and novel sources of data but the evidence to date suggests that just as the fundamental principles of social science methodology will hold in this new research environment so too will classic principles of archiving and managing collections of research data for secondary data analysis.

The core strength of classification schemes such as UDC is that a classification code crosses all borders. The code is not only instantly recognisable but also, wherever it is used to classify research data, the opportunity is there to share subject metadata.

References

- Barbalet, S. (2015) Enhancing Subject Authority Control at the UK Data Archive: A Pilot Study Using UDC. In Slavic, A. & Cordeiro, M. I. *Classification and Authority Control. Proceedings of the International UDC Seminar*. Ergon-Verlag
- Borne, K. (2016) Rocket-Powered Data Science: Data Reflections. <http://rocketdatascience.org/> Accessed 24/02/2017.
- Broughton, V. (2015) *Essential Cataloguing*. 2nd ed. London, Facet.
- Corti L. et. al. (2014) *Managing and Sharing Research Data: A Guide to Good Practice*. Los Angeles: Sage.
- Hjorland, B. (2017) Subject (of Documents) *Knowledge Organisation* 44 (1): 55-64.
- Lazer, D. et. al. (2014) The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 14 Mar, 343 (Issue 6176): 1203-1205 <http://science.sciencemag.org/content/343/6176/1203.full> Accessed 27/02/2017
- Slavic, A. (2006) UDC In Subject Gateways: Experiment or Opportunity? *Knowledge Organization* 33 (2): 67-85.
- Tay, A. (2016) Managing Volume in Discovery Systems. In Spiteri, L ed. *Managing Metadata in Web-Scale Discovery Systems*. London, Facet.

⁵ See <https://www.theguardian.com/technology/2014/mar/27/google-flu-trends-predicting-flu>

Providing full and equal access to resources is a primary responsibility of libraries in multi-cultural societies, but when these cultures find expression in multiple languages and scripts, libraries can find themselves stymied by the limits of technology and international standards.

Israel is one such multi-cultural, multi-lingual hotbed, with two official languages (Hebrew and Arabic), one commonly accepted language of academia (English), and, as a nation with a high percentage of immigrants, several smaller language groups. The primary scripts in use are Hebrew, Arabic, Latin, and Cyrillic. No one language is the native tongue of a majority of Israelis – only 49% of Israelis over the age of 20 are native Hebrew speakers, and a further 18% have Arabic as their native tongue and 15% Russian (Central Bureau of Statistics, 2013). A solution such as practiced abroad – romanization – is unworkable for access by the general public, both because of the system’s emphasis on grammatical purity over relevance to spoken language and because it uses an alphabet which is not native to the vast majority of Israelis.

Cutter’s (1904, p. 12) objectives for a dictionary catalogue hold that the function of the catalogue is to show what the library has:

- (a) by a given author
- (b) on a given subject
- (c) in a given kind of literature.

With resources available in nearly every language and four languages (Arabic, English, Hebrew, and Russian) used in cataloguing, how could Israeli libraries provide access to the breadth and width of their collections? The traditional answer, which was used in the era of card catalogues, was having separate catalogues for each script. To find all resources on a topic or by a creator the patron had to run four separate searches. This is inefficient, unwieldy, and utterly unsuited to the digital age.

Libraries have tried various technological solutions to overcome this difficulty. The most common work-around is to give an added entry for names in additional scripts. While this allows for discovery in multiple scripts it distorts the bibliographic record by recording the same entity twice (the “phantom tracing” warned against by Malinconico in 1974 [1985]) and does not solve the problem of simultaneous searching across scripts. Added entries cannot be updated automatically from unilingual authority records, so cataloguing staff must do more manual bibliographic file management. This system also places an added cognitive burden on patrons, who need to learn that added entries sometimes refer to separate agents and sometimes to the same, requiring them to distinguish between the two cases and know when they must follow up on the added entry to see all of a creator’s work.

In any case, added entries for agents already named in the record leads to the number of resources in the catalogue being inconsistent, since the same resource is counted in each language, leading to duplicate mentions of the same manifestation.

רשימת חיפוש לפי: Authors+Titles		רשימת חיפוש לפי: מחברים	
מספר רשומות	ערך	מספר רשומות	ערך
2	Kedar, Rochelle	10	קדר, רחל
1	Kedar, Shelley	5	קדר, רחל (תדפיסים)
1	Kedar, Yig'al	2	קדר, רמי
1	Keddie, Kenneth M. G	1	קדר, רנה - רמיזות
1	Keddie, Nell	1	קדר, רקפת
17	Keddie, Nikki R	2	קדר, שאול
1	Keddie, Nikki R Roots of revolution	3	קדר, שדמה
1	Kedem, Benjamin	1	קדר, שלי - רמיזות
2	Kedem-Cuomo, Dina		קדר, שמואל - רמיזות
1	Kedem, Michal		ראה: קדר, שמואל אליקים
		30	קדר, שמואל אליקים - רמיזות

Fig. 1: The goal of full access is achieved, but not of equal access. Two of the 10 resources attributed to the Hebrew form of Kedar, Rochelle are English-language expressions. Since the Hebrew form was added to those cataloguing records, they show up in both the Hebrew-language form and the English. Users would be justified in thinking the author created 12 manifestations: 2 in English and 10 in Hebrew.

The National Library Law (2007) made the National Library of Israel (NLI) independent of its academic parent body. Part of the new mandate was “to make the Library’s collections reasonably accessible to the general public, in Israel and abroad, also through advanced technologies.” The lack of full and equal access to NLI’s multilingual collection went from being an annoyance to being a potential legal problem; the search for a solution was on.

Though many European libraries have multilingual collections, they are, primarily, in a single script. Therefore they were of limited help in deriving a solution. The libraries of the Arabic-speaking world experience problems similar to Israel’s, but they, too, had no answers which could be adopted.

Though the Bibliotheca Alexandrina uses “the Library of Congress MARC 21 Model B, where the transcribed text in the bibliographic record is entered only in the script in which it appears” (El-Sherbini, 2013), their authority records are uniscript. The same agent may have multiple name authority records and the problem of simultaneous access is not solved.

In 2011 NLI ramped up efforts to solve the problem. The strongest contender for a solution was using the Virtual International Authority File (VIAF) to create a linked data solution, but at that point VIAF was not yet considered a stable product. In any case, using VIAF as the mechanism to link various forms of an agent’s name would mean losing all notes and justifications in the local database. Using VIAF would have also meant giving up local control, being unsure what sort of results patrons would get, and would have involved a lot of manual bibliographic file maintenance as VIAF clusters changed. Most importantly, existing library systems could not make sophisticated enough use of VIAF data to give patrons a unified and uniform search experience.

The problem remained: How could NLI, and in its wake other Israeli libraries, allow all resources to be discovered in a single, multilingual vernacular search? The answer: NNL10, a multilingual, multiscript authority database that combines MARC encoding with local subfields to create robust equivalent terms.

NNL10 started its life as NLI's authority file. It contains two major divisions:

- name authorities (people, corporate bodies, families, titles, series)
- subjects (Library of Congress Subject Headings [LCSH], translations of LCSH, localized subjects, geographical entities)

In keeping with NLI's initial research, VIAF clustering was used to bring together the original unilingual authority records. The Swiss National Library's use of multiple 1XX fields with a local subfield (\$9) to separate authorized access points in different languages (all in one script) was adopted to bring together the four legal scripts of authorized and variant access points in a single authority file.

<u>010</u>	—	a	n 50051544	
<u>040</u>	—	b	mul	
		e	rda	
<u>046</u>	—	f	1939-05-04	
		z	edtf	
<u>100</u>	<u>1</u>	a	Oz, Amos,	
		d	1939-	
		l	lat	
<u>100</u>	<u>1</u>	a	Oz, Amoc,	
		d	1939-	
		l	cyr	
<u>100</u>	<u>1</u>	a		עוז, עמוס,
		d		1939-
		l		heb
<u>100</u>	<u>1</u>	a		عوز, عاموس,
		d		1939-
		l		ara
<u>370</u>	—	a	Jerusalem (Israel)	
<u>374</u>	—	a	Author	
<u>375</u>	—	a	male	
<u>400</u>	<u>1</u>	a	Ouzi,	
		d	1939-	
		l	lat	
<u>400</u>	<u>1</u>	a	Ao, Zi,	
		d	1939-	
		l	lat	
<u>400</u>	<u>1</u>	a	Klausner, Amos	
		d	lat	
<u>400</u>	<u>1</u>	a	Ozs, Amoss	
		d	lat	
<u>400</u>	<u>1</u>	a	Os, Amos	
		d	lat	
<u>400</u>	<u>1</u>	a	Oz, Amosz	
		d	lat	
<u>400</u>	<u>1</u>	a	Klozner, Amos	
		d	lat	

Fig. 2: Part of the NNL10 record for Israeli author Amos Oz, showing four instances of the 100 field and use of the local subfield \$9 to distinguish between scripts.

Along with information from VIAF, Israeli databases were mined for access points and attributes. Among these sources: the Bibliography of the Hebrew Book, Imagine (database of Israeli artists), and ACUM (database of Israeli musicians).

Latin character authorized access points match Library of Congress records and Arabic and Cyrillic access points are designed to match the Latin character access point as closely as possible. Hebrew access points, on the other hand, follow the agent's preferred spelling or the best of local scholarship. Therefore there may be an imperfect match between the four authorized access points representing a single agent, but each will be the access point most recognizable to the user.

The other side of NNL10 is multilingual access LCSH. At present available in Hebrew and English, NNL10 offers the same full and equal access to subjects as it does to name authorities. In addition, NNL10 provides a translation of every subject and subdivision, precoordinates millions of subject headings, and has tables governing which combinations are legal and warnings against those which are not. Indexers need not study the complex rules of the Subject Heading Manual to accurately construct compliant subject headings.

NNL10 provides an answer to a question which *“has been recognized as an issue to be solved by the Association for Library Collections and Technical Services (ALCTS) Task Force on non-English Access as far back as 2007”* (El-Sherbini, 2016).

The problem with using LCSH for a multilingual database is well-known: LCSH reflects the American culture, vocabulary, and law, while a multilingual society is likely to be multi-cultural. In the interest of standardization it was decided to retain American decisions about general subject headings, but as in the name authority records, it was decided that Israeli headings would be subject to adaptation. Jerusalem, Israel’s capitol, is not left stateless in NNL10 as in the Library of Congress database. Israeli place names were determined by an Israeli government committee; the decision on whether to qualify a geographical location as Israel, West Bank, or Gaza Strip was based on official Israeli, and not United States, governmental policy. The pre-State geographic area which is now Israel is given the subdivision “Eretz Israel,” the Jewish term for the area, rather than the Greek term “Palestine”. Local conflicts and wars are given their Israeli names; no matter what their native language, no Israeli would think of the Six Day War as “Israel-Arab War, 1967”.

After creating authority files for local subject headings, existing partial translations of the remaining LCSH terms were imported. These translations were made over the past two decades by two Israeli universities and followed two different orthographic principles – defective and plene spelling. Batch conversion routines changed nearly all defective spelling to plene, and crowdsourced quality control is quickly finding the outliers to mark for manual correction.

Though this effort is expected from a national library, it was felt that the authority database would only truly fulfill its potential if it became a cooperative effort; not only would libraries use the information as the basis of their own local authority files, but they would install NNL10 in place of their local files. Thus, patrons searching the national union list of bibliographic records would enjoy the same full and equal access to names and subjects as they do when searching NLI’s catalogue.

As a result of outreach efforts, in 2014 two major universities, Haifa University and the Hebrew University of Jerusalem, joined the NNL10 cooperative. In 2015 four colleges (Beit Berl, David Yellin, Kinneret, and Oranim) joined, and in 2016 three more universities, Ben Gurion University of the Negev, Bar Ilan University, and Tel Aviv University, began the switch from local authority records to NNL10.

Libraries which join have a local copy of NNL10 uploaded to their servers. An automated routine written by NLI runs on the local catalogue and adds the appropriate \$9 subfield to all authority controlled fields in the bibliographic records and to 1XX, 4XX, and 5XX fields in the authority records. Another routine identifies overlaps in 1XX or 4XX between the local catalogue and NNL10. These overlaps are then converted, in the bibliographic record, to the NNL10 authorized access point in the script of the bibliographic record.

The percentage of out-of-the-box matches between local catalogues and NNL10 depends on the quality of each library’s authority work and its adherence to national and international cataloguing standards. To create more matches additional routines are run; one highly successful routine compares 1XX which are similar but not exactly the same (a name with and without dates, for example) and creates a match if the two headings share bibliographic works.

Partners receive individualized training and then are authorized to create new 1XX, 4XX, and 5XX fields for name authority records. They may not change or merge existing 1XX fields but must request such changes from NLI cataloguing department managers. Only NLI is allowed to make changes to subject heading authority files.

Working with the various institutions, NLI has managed to overcome compatibility issues between different versions of Ex Libris's Aleph program and to solve the problem of labelling authorized access points in Aleph authority records to automatically update bibliographic records in Ex Libris's Alma program via the Community Zone, thus allowing NNL10 partners who have switched to Alma to completely dispense with their own local authority files.

With over two thirds of Israeli universities now partners in NNL10, NLI decided to make the database publically available through the national union list (http://uli.nli.org.il/F/7B9BTLNCJD6RVVU4HSK79SFV8GPM9KTC1GQE8N2B6JPSN536MF-35582?func=find-b-0&local_base=nnl10&con_lng=eng). The cooperative is in the midst of expanding the database (over a quarter million name authorities are expected to be added over 2017) and NLI is currently working on plans to release both name and subject information as linked data.

References:

- Central Bureau of Statistics. (2013). Selected Data from the 2011 Social Survey on Mastery of the Hebrew Language and Usage of Languages. Retrieved November 11, 2016, from http://www.cbs.gov.il/reader/newhodaot/hodaa_template.html?hodaa=201319017
- Cutter, C. A. (1904). *Rules for a dictionary catalog* (4th ed.). Washington: Government Printing Office. Retrieved from <https://ia800205.us.archive.org/13/items/rulesforadictio06cuttgoog/rulesforadictio06cuttgoog.pdf>
- El-Sherbini, M. (2013). Biblioteka Alexandrina's model for Arabic name authority control. *Library Resources & Technical Services*, 57(1): 4–17.
- El-Sherbini, M. (2016). Improving resource discoverability for non-Roman language collections. In *IFLA WLIC 2016, 11-12 August 2016, Columbus, Ohio in Session Subject Access: Unlimited Opportunities*.
- Malinconico, S. M. (1985). The role of a machine based authority file in an automated bibliographic system. In M. Carpenter & E. Svenonius (Eds.), *Foundations of cataloging: A sourcebook* (pp. 211–233). Littleton, Colorado: Libraries Unlimited.
- The National Library Law (2007). Retrieved from http://web.nli.org.il/sites/NLI/English/library/aboutus/renewal/Documents/National_Library_Law.pdf

Introduction

This short communication summarises some of the ideas shared at CIG 2016 (Welsh 2016b, Welsh, 2016c) regarding the current state of art in the bibliographic model used by the Anglo-American cataloguing community. It highlights our trajectory from the ‘monolithic’ catalogue record of the card and dictionary catalogue era, through the hyperlinked computer-based records of the 20th century, towards models that focus on data that can be reused and repurposed within the 21st century linked data environment.

The Four Models

As presented elsewhere (Welsh et al, 2016), in order to fully understand modern cataloguing and meet the needs of employers, cataloguers need to have a working awareness of the four bibliographic models utilised in the Anglo-American cataloguing community (Table 1).

Card / Dictionary Catalogue	Paper-based system; entry points and length of description limited by concerns around physical space (size of card / page and size of catalogue cabinet / shelving unit). Each record describes the item in the cataloguer’s hand entirely – its intellectual and physical contents.
ISBD and AACR in MARC	ISBD, AACR and AACR2; built on principles of the card and dictionary catalogue; monolithic record structure in which information in fields is contextualised and explained by reading the entire record; inherited the card / dictionary catalogue’s description level, covering both the intellectual and physical contents of the item in hand.
FRBR in RDA	Although much is inherited from ISBD, AACR and AACR2, the principle of RDA is a version of FRBR’s WEMI model: Works, Expressions, Manifestations and Items are catalogued separately and linked to each other (and to the appropriate WEMI of other records); moves away from the monolithic record, favouring instead field-level links between records, aiming for the linked data environment. Currently constricted by systems based on the MARC format, which does not accommodate RDA’s linked data capacities.
BIBFRAME	Linked data solution proposed by the Library of Congress; based on RDF; data model proposes Work, Instance, and Item, allowing for simplified analysis on the part of cataloguers and, significantly, the construction of RDF triples – a foundation for linked data in the wider web environment. Many triples in BIBFRAME incorporate Annotations – pieces of data that provide information about the Work or Manifestation, such as creator, publisher or ISBN.

Table 1. Cataloguing Models Anglo-American cataloguers learn in 2016, updated from Welsh et al, 2015 to reflect BIBFRAME 2.0

Given the uptake of *Resource Description and Access* (RDA) in the UK, cataloguers are now quite familiar with its implementation of the FRBR model, and even although it cannot be fully exploited in MARC, its concepts of the Work which is realized through an Expression which is embodied in a Manifestation which is exemplified in an Item have fed through into our cataloguer jargon, for better or for worse. Figure 1 presents an entity relationship diagram for Walter de la Mare's (1930) *Desert Islands and Robinson Crusoe* now in the collection of Cambridge University Library (Classmark CCB.49.1), a WEMI example originally discussed in an article in the National Acquisition Group's journal *Taking Stock* (Welsh and Wright, 2016).

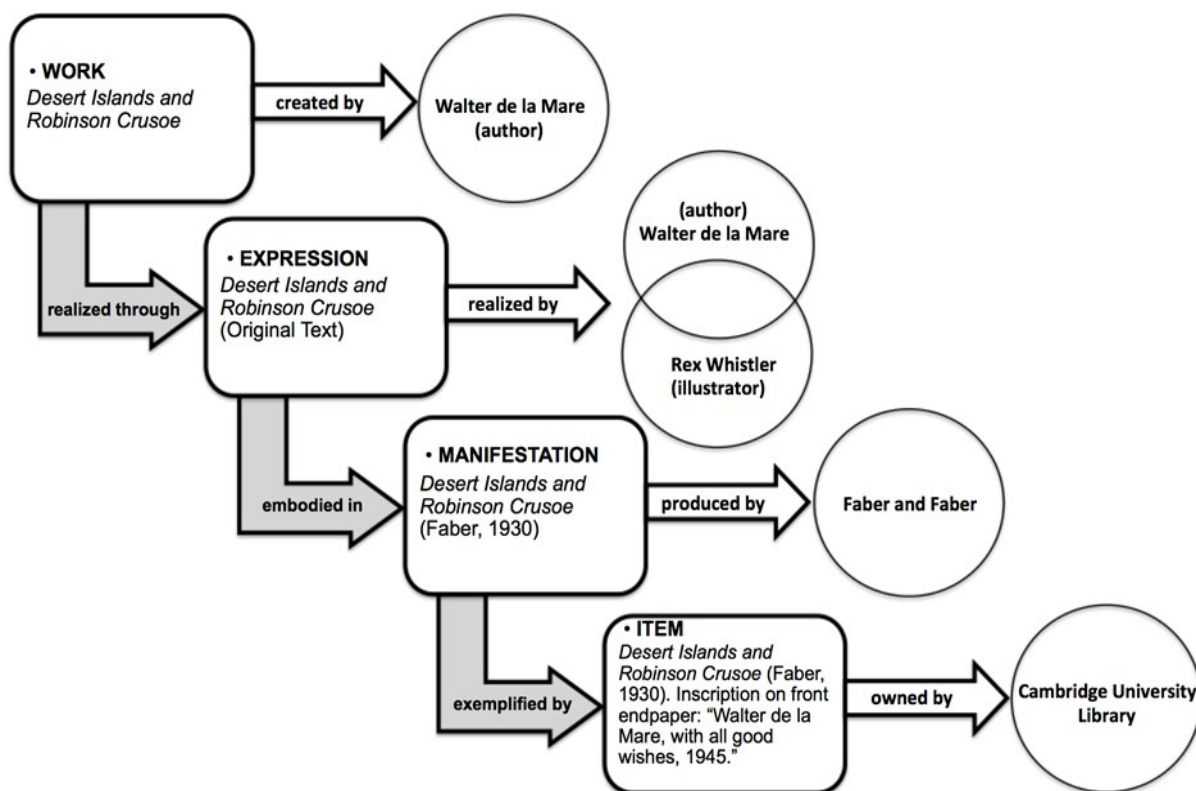


Figure 1. Group 1 and 2 entities and relationships for *Desert Islands and Robinson Crusoe* (Faber, 1930), Cambridge University Library CCB.49.1, originally published in Welsh and Wright, 2016.

The diagram demonstrates the most common relationships between “Group 1 entities” (Works, Expressions, Manifestations and Items) and “Group 2 entities” (people, corporate bodies and families):

- A Work is created by a person, corporate body or family (in this case, the person Walter de la Mare)
- An Expression is realized by a person, corporate body or family (in this case the persons Walter de la Mare and Rex Whistler)
- A Manifestation is produced by a person, corporate body or family (in this case the corporate body Faber and Faber)
- An Item is owned by a person, corporate body or family (in this case Cambridge University Library).

Relationships at Heart

In a bibliographic model which places “relationships at the heart of the catalogue” (Welsh and Batley, 2012, 8), thinking about the people or organisations responsible for what we are describing can help us to gain a firmer idea in our own minds of which Group 1 entity we have.

In this example, Rex Whistler was commissioned to illustrate de la Mare's work, and so his involvement was in its realization – therefore, at the Expression level. However there are times in which a writer and an artist work together to create a Work, or in which the creators are both artists and writers. Figure 2 shows the WEMI relationships for *A Child of Books* (Jeffers and Winston, 2016), whose co-creator Sam Winston was our after-dinner speaker at CIG 2016.

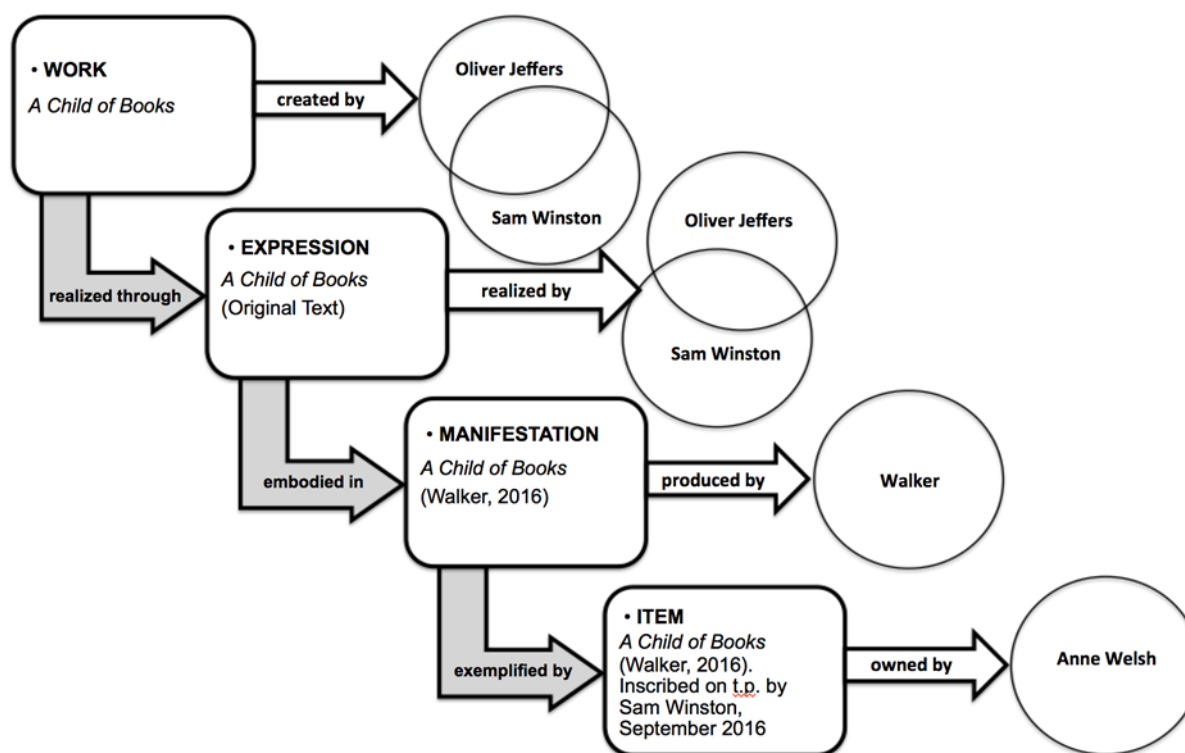


Figure 2. Group 1 and 2 entities and relationships for *A Child of Books* by Oliver Jeffers and Sam Winston, inscribed by Sam Winston to Anne Welsh.

As cataloguers, we have to take what appears on the title page as our chief source of information, and in the case of *A Child of Books*, this is easy:

245 12 \$aA child of books /\$cOliver Jeffers, Sam Winston

Similarly for *Desert Islands and Robinson Crusoe*:

245 10 \$aDesert islands and Robinson Crusoe /\$cby Walter de la Mare ; with decorations by Rex Whistler

So, in the purely practical terms of what we “do” with the statement of responsibility, both are equally straightforward. However, when we think about the WEMI, there are clear differences. Rex Whistler was a well-known artist and book illustrator, and his work with de la Mare came under his practice in illustration – that is taking the words that someone else (in this case, de la Mare) had written already and providing artwork to accompany them. We could, of course, go through the publisher archives to discover the exact nature of the contribution of each person, but, at base level, the Work consists of a text by de la Mare, and Whistler's decorations form part of its Expression. We can think of other Works that exist in this way – Michael Bond's character Paddington Bear was inspired by a teddy bear he bought for his then wife, famously illustrated by Peggy Fortnum in the original books in the 1950s and 1960s, then by Fred Bandery in the Young Readers series from the 1970s, before being transformed into a cartoon by Ivor Wood, who also created the puppet used in the 70s TV series.

Later illustrators include David McKee and Barry Macey in the 1980s and RW Alley from the 1990s. For most of the publications featuring Paddington, at Work level Michael Bond is responsible, and then the illustrators become involved in its realization through different Expressions. Paddington could, in fact, be the subject of an entire workshop and article on his own. For those interested in a quick summary of the famous bear's illustrated evolution, *The Guardian* shared a gallery of artwork to celebrate Bond's 90th birthday in January 2016 ('Paddington through the Ages,' 2016).

Relationships between People, not Solely Resources

Jeffers's and Winston's collaboration is interesting for artistic reasons, but also because it is well documented through interviews with them. As an article in *The Telegraph* (Calkin, 2016) outlined, "When Jeffers and Winston were introduced ... they decided to collaborate, despite – or perhaps because of – their different styles. They recognised something in each other: both have a love of literature rendered as art." Jeffers is quoted as saying, "I thought something interesting would happen if we put my drawings together with Sam's landscapes," while Winston is reported as sharing that, "It was a dreamy kind of idea to start with ... It could be a book, a sculpture, a set of paintings ... we were playing around - on the shoreline where ideas come from."

Here, we can really see that in thinking about how the bibliographic entities of WEMI and creators relate to each other, we are not solely thinking as people tasked with data entry for a book. We are also engaging with ideas around the relationships between people and the intellectual and artistic outputs they are creating. As we move "down" the WEMI model through the Manifestation level, we are also thinking about and documenting the relationship between the book and its publisher, and, through the book, between the creators and *their* publisher – in this case, Jeffers's and Winston's Work is realized *by them* in its Expression, which is embodied in a Manifestation produced by *their publisher, Walker*. In the final level of WEMI, another relationship is represented – Jeffers's, Winston's and Walker's Manifestation is exemplified by an Item that is owned by me.

Because there is an inscription, my relationship to the book is documented: it is clear that I have met Sam Winston at some point because he has inscribed the title page, whereas Oliver Jeffers has not. There are limits to the level of revelation this provenance evidence provides. There is a date, but nothing to indicate what we know at CIG – that Sam Winston signed the book at our conference, at which he spoke after dinner and where Walker had a stall. Nor can there be anything to indicate that Karen Pierce and I had, in fact, met Winston when he provided a keynote at the Livres d'artistes conference in Cardiff the previous December and that we suggested to the CIG conference committee, of which Karen was part, that they might invite him to speak after our conference dinner. However, the recording of the basic provenance information could set a scholar on the road to such a discovery, should anyone ever wish to trace, for example, the history of the Cataloguing and Indexing Group conferences in the 20th and 21st centuries – a suggestion that does not seem as flippant as it first sounds when I look at some of the work I am currently doing around the institutional history of UCL Department of Information Studies as it approaches its centenary in 2019.

Here we can see that final relationship, chronologically, between the researcher and the materials they seek – often in pursuit of information about a literary or artistic creator – is, as argued vociferously elsewhere (Welsh, 2016d), facilitated by the intellectual work of the cataloguer.

From Records to Data

In recording information about relationships between people and resources and through this between people and other people, we are meeting the challenge of going beyond what FRBR has called "the elementary uses that are made of the data by the user" (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998). In establishing intellectual links of the kind WEMI forces us to consider, we are also moving beyond what Wilson (1968) termed "descriptive power" to "exploitative power," defined by Smiraglia (2008) as "the power of a scholar to make the best possible use of recorded knowledge,"

and enabling us to meet what I have described as “the needs of researchers who want to engage with our data not [solely] as a route through to ‘the real’ objects of their research – full-text files, books, the item for which the catalog data is a surrogate – but as an integral part of their own research” (Welsh, 2016d).

Linked data is powered by relationships. As Godby and Smith-Yoshimura (2017, 18) put it in their recent practical article on ‘Managing the Transition from Legacy Library Metadata to Linked Data,’ “To make linked data work, the library community needs good data that is structured, unambiguous and published in a format that enables linking with data produced by other communities. Library data also needs to be more about the *Things* or the people, organizations, places and topics that users care about and the library community has something to say about” (authors’ emphasis). In short, in entering the linked data environment, we are moving from the monolithic catalogue record, in which most fields and sub-fields are human-readable and dependent on being read within the context of the record as a whole, to models in which we present our data in ways that can be combined and recombined; in which each field is not dependent on another field in order to be understood; and in which many of the key pieces of data are not human readable, but machine-readable. In the Internet of Things, each Thing has a Unique Resource Identifier (URI) to which those who wish to reuse that data can point using semantic web specifications such as RDF (the Resource Description Framework) in order to describe relationships between that Thing and any other Thing with a URI. RDF is formed using ‘triples’ in the format shown in Figure 3.

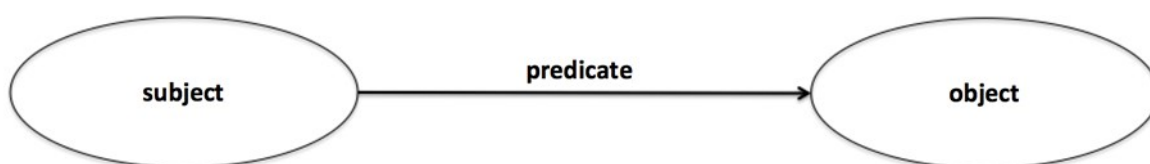


Figure 3. Standard format of a triple in RDF

BIBFRAME Triples

The format being developed by the Library of Congress to replace MARC is BIBFRAME (the Bibliographic Framework Initiative). Built on RDF, it describes relationships between data using RDF triples in the format exemplified in Figure 4. Where RDA follows FRBR in using the WEMI model, BIBFRAME has a simplified version, based initially on only two levels, Work and Instance, and from April 2016, when BIBFRAME 2.0 was announced, three levels, Work, Instance and Item. The easiest way to visualize the mapping between the FRBR in RDA and BIBFRAME models is to think of their Work and Item levels as being equivalent and the Instance level of BIBFRAME equating to both Expression and Manifestation levels in FRBR in RDA. Figure 4 shows the BIBFRAME triple that links data at the Work level to data at the Instance level. Figures 5 and 6 show this triple ‘in action’ with regard to *Desert Islands and Robinson Crusoe* and *A Child of Books*.



Figure 4. Example of a BIBFRAME triple showing the subject (WORK) with predicate (hasInstance) and object (INSTANCE)



Figure 5. Example of a BIBFRAME triple in which the subject is the Work *Desert Islands and Robinson Crusoe* and has the predicate “hasInstance” and the object is the Instance published by Faber in 1930.



Figure 6. Example of a BIBFRAME triple in which the subject is the Work *A Child of Books* has the predicate “hasInstance” and the object is the Instance published by Walker in 2016.

Using triples defined by BIBFRAME, it is possible to model a web of bibliographic data, in which we can map the relationships between people and resources. One way to create such a web quickly and easily with minimum use of technology is to work in small groups and build it from post-it notes. Figure 7 shows part of such an exercise undertaken by the Advanced Cataloguing class at UCL in February 2017, using the same Jane Austen and Fay Weldon examples we used in workshops at CIG 2016. The UCL example is being shown in preference to the CIG one only because the post-its I was able to acquire were narrow and so could indicate a vague form of direction and had different coloured edges allowing for a loose form of colour coding, so that even though it is difficult to read the writing in the photograph here, it is possible to gain a sense of the web of connections that is forming.

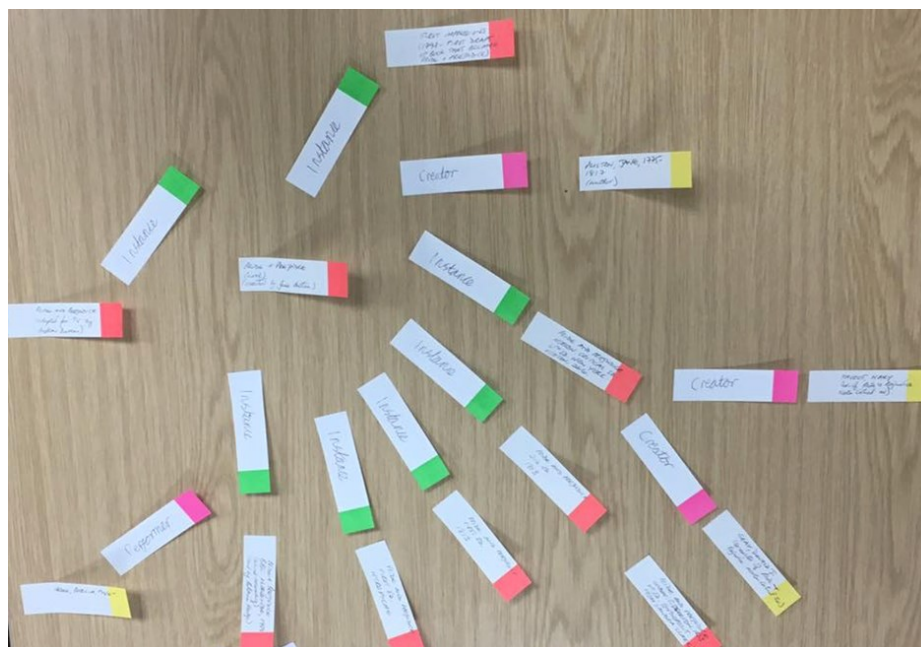


Figure 7. Post-its arranged by students in INSTG005 Advanced Cataloguing in 2017 to plot relationships between some Works, Expressions, Manifestations and Items by Jane Austen and Fay Weldon.

From *Pride and Prejudice* to *The Cloning of Joanna May*

I wrote examples of BIBFRAME Works, Instances and Items on the orange-tipped post-its. The students used the green-tipped post-its to indicate Instance relationships (BIBFRAME's `hasInstance` predicate) linking these entities. I wrote the names of people involved in the creation or production of the Works, Instances and Items on the yellow-tipped post-its and the students indicated the type of relationship (e.g. BIBFRAME's `creator`) on the pink-tipped post-its. The post-it diagram we created covered two small desks, and Figure 7 shows an excerpt from it.

Having created the diagram, it was possible for students (and participants at CIG 2016 before them) to see how one could trace a journey through different relationships from the Work *Pride and Prejudice* to the unrelated Work *The Cloning of Joanna May*. In the set of data I had provided on the post-its, the point at which a link appeared was through the actor Patricia Hodge, who read both the HarperCollins Audio recording of *The Cloning of Joanna May* (Weldon, 1996) and the BBC Worldwide recording of *Pride and Prejudice* (Austen, 1999). Clearly, there is no direct relationship between the two Works, but it is fairly obvious how BIBFRAME (or another RDF schema) allows us to plot relationships between people and objects that ultimately do form links. It is also clear that someone looking for material could enter the web we created at different points and still go on a journey – they might, for example, come in with a specific known item search for the Norton Critical Edition of *Pride and Prejudice* (Austen, 2016) and leave having also discovered by following the linked data that Austen's first draft of the book was actually called *First Impressions* (Austen, 1797). Having been reputedly disposed of by the publisher Caddell to whom Austen's father had sent it, the 1797 manuscript of *First Impressions* is of interest within our bibliographic model in that it exists only as a Work that is related to the Work *Pride and Prejudice*. It does not have an Instance, nor, indeed, an Item.

Items Matter

One question that I anticipated being asked by colleagues at CIG 2016 who have been following the development of BIBFRAME, and that I was indeed asked, was why BIBFRAME 2.0 was needed and what impact it might have on our practice. My poster on 'Provenance in the Bibliographic Model' (Welsh, 2016c) considered one of the major improvements made to BIBFRAME by the new version released in April – the introduction of the Item level and the predicate `custodialHistory` that can be used to form a triple in the form shown in Figure 8.



Figure 8. Example of a BIBFRAME triple showing the subject (ITEM) with predicate (`custodialHistory`) and object (custodian).

Figure 9 shows this triple for Cambridge University Library's copy of *Desert Islands and Robinson Crusoe*.



Figure 9. Example of a BIBFRAME triple showing the subject (the Item *Desert Islands and Robinson Crusoe* (Faber, 1930) inscribed on front endpaper: "Walter de la Mare, with all good wishes, 1945") with predicate (`custodialHistory`) and object (Cambridge University Library).

Figure 10 shows this triple within a web of some of the triples that could be created in BIBFRAME in order to describe the resource. You can see that in BIBFRAME both the author (de la Mare) and the illustrator (Whistler) are indicated by “agent” but that de la Mare’s relationship is at the Work level, while Whistler’s is at the Instance level.

The significance of the introduction of BIBFRAME’s *custodialHistory* relationship is that it should be possible to record the chain of custody of objects within collections. Figure 11 shows how it might be possible to show the known history of the Cambridge University Library copy of *Desert Islands and Robinson Crusoe*. The copy clearly started in the hands of de la Mare, as indicated by his inscription. At this point, we don’t know if it was a gift to Phyllis T.M. Davies or to another person and Davies acquired it through purchase, but we do know that she was the person who owned it before Cambridge University Library obtained it as part of her collection, and we know that the copy is still sitting in the tower at the UL.

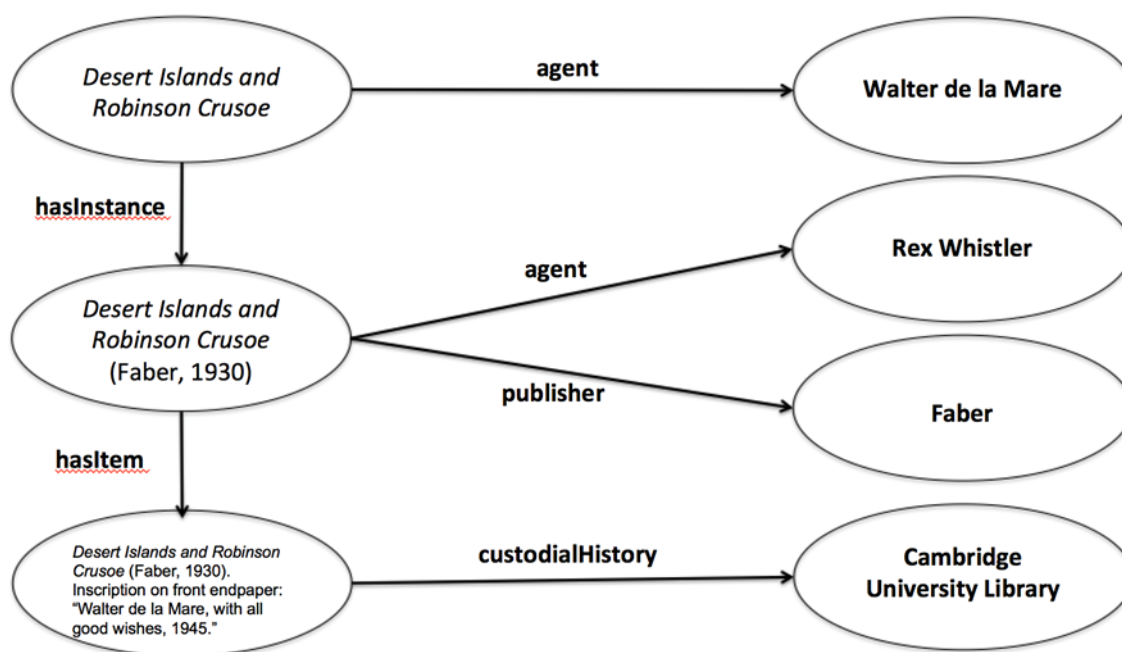


Figure 10. Example of BIBFRAME triples showing some of the relationships we can trace for Cambridge University Library’s copy of *Desert Islands and Robinson Crusoe*.

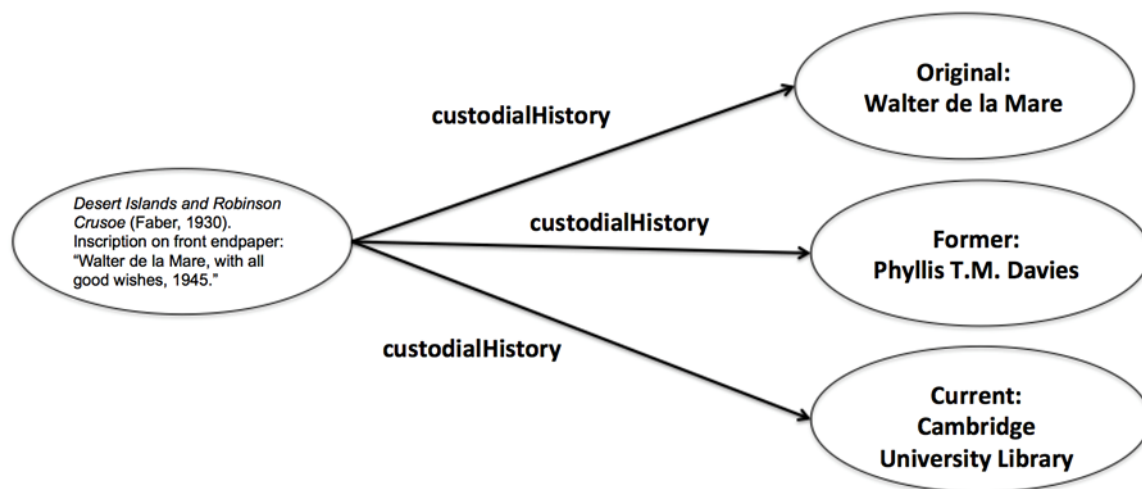


Figure 11. Example of BIBFRAME triples showing some of the chain of custody for Cambridge University Library's copy of *Desert Islands and Robinson Crusoe*.

From this it should be clear that the Library of Congress in creating BIBFRAME is attempting to position library data not only in a way that is more compatible with other forms of heritage data in the linked environment, but that allows cataloguers to record information that we know to be of interest to researcher (such as provenance information) in a way that will be easy to reuse. This element of cataloguing is contributing to the history of the objects within our collections and helping researchers to make connections quickly and easily, based on our intellectual efforts. In this way, a cataloguer can be seen as a “distant research collaborator” (Welsh, 2016d), assisting in the task Wilson (1968) highlights as “exploitative power,” making “the best possible use of recorded knowledge” (Smiraglia, 2008).

For current cataloguers, there is no rush to implement these ideas. BIBFRAME is still evolving, and, therefore, is some way from being introduced within the Library Management Systems that we use to create and manage our data. At this stage, the need to engage with these ideas is preparatory – it's a good idea for us to be aware of the changes we will manage in the next few years, and, as always, it is helpful to be able to contribute to discussions with our management teams, our library vendors and with those who are creating the standards, rules and, in the case of BIBFRAME, the exchange formats that we will be expected to use. While computer scientists know more about data models and theorists can expound on the Book (and other formats) and its fundamental role and relationships, it is the cataloguer sitting quietly at their desk creating data day in and day out who knows better than anyone the full range of quirks and joyous oddities that challenge not only their own decision making skills, but the full capacity of our standards and formats created at the highest levels internationally. It is the everyday cataloguer who wields Wilson's (1968) “two kinds of power” – “descriptive” and, increasingly, “exploitative.”

Acknowledgements

The author would like to thank participants in the workshop at CIG 2016 on ‘Getting to Grips with Bibliographic Models’ and UCL's INSTG005 Advanced Cataloguing class for highlighting the things that they found challenging and interesting in the workshops, which in turn focused this article. The example from The Phyllis T.M. Davies Collection at Cambridge University Library was discovered while undertaking research for a PhD in Cultural Studies at University College London.

Works Cited

- Austen, J. (1797) *First Impressions*. Unpublished manuscript.
- Austen, J. (1999) *Pride and Prejudice*. Sound recording. (London: BBC Worldwide).
- Austen, J. (2016) *Pride and Prejudice*. Norton Critical Edition edited by Donald, D.J and Favret, M.A. (New York: Norton).
- Calkin, J. (2016) 'A Child of Books: Making its Mark in Children's Literature.' *The Telegraph* (1 September), <http://www.telegraph.co.uk/books/childrens-books/a-child-of-books-making-its-mark-in-childrens-literature/>
- de la Mare, W. (1930) *Desert Islands and Robinson Crusoe* (London: Faber).
- Godby, C.J and Smith-Yoshimura, K. (2017) 'From Records to Things: Managing the Transition from Legacy Library Metadata to Linked Data.' *Bulletin of The Association for Information Science and Technology* 43(2): 18-23.
- IFLA Study Group on the Functional Requirements for Bibliographic Records (1998) *Functional Requirements for Bibliographic Records Final Report* (Munich: K.G. Saur).
- Jeffers, O. and Winston, S. (2016) *A Child of Books* (London: Walker).
- 'Paddington through the Ages: in Pictures.' (2016) *The Guardian* (13 January), <https://www.theguardian.com/childrens-books-site/gallery/2016/jan/13/paddington-through-the-ages-michael-bond>
- Smiraglia, R.P. (2008) 'Rethinking What We Catalog: Documents as Cultural Artifacts'. *Cataloging and Classification Quarterly* 45: 25-37.
- Weldon, F. (1996) *The Cloning of Joanna May*. Sound recording. (London: HarperCollins Audio).
- Welsh, A. (2016a) 'Cataloguer as Distant Research Collaborator: Implications of the Use of Catalogue Data in Humanities Research.' Presented at: Emerging Research in Library and Information Science, London, 19 August 2016, <https://iris.ucl.ac.uk/iris/publication/1154530/1>
- Welsh, A. (2016b) 'Getting to Grips with Bibliographic Models.' Presented at: Cataloguing and Indexing Group Conference, Swansea, 31 August – 2 September 2016, <http://www.cilip.org.uk/cataloguing-indexing-group/presentations/conference-2016-innovation-discovery>
- Welsh, A. (2016c) 'Provenance in the Bibliographic Model.' Presented at: Cataloguing and Indexing Group Conference, Swansea, 31 August – 2 September 2016, <http://www.cilip.org.uk/cataloguing-indexing-group/presentations/conference-2016-innovation-discovery>
- Welsh, A. (2016d) 'The Rare Books Catalog and the Scholarly Database.' *Cataloging and Classification Quarterly* 54(5-6): 317-337.
- Welsh, A., Bikakis, A, Garea Garcia, N., Mahony, S., Inskip, C. and Vogel, M. (2015) 'Work in Progress: The Linked Open Bibliographic Data Project.' *Catalogue and Index* 178: 16-21.
- Welsh, A. and Batley, S. (2012) *Practical Cataloguing: AACR, RDA and MARC 21* (London: Facet).
- Welsh, A. and Wright, J. (2016) 'RDA for Acquisitions Staff.' *Taking Stock* 24(2): 14-16.
- Wilson, P. (1968) *Two Kinds of Power: An Essay on Bibliographical Control*. Berkeley: University of California Press.

Introduction

I was recently given access to Google Analytics for the purpose of analysing usage of our discovery layer and our library management system provided catalogue. I was looking for a way to assess how users interact with the search systems we provide and whether the assumptions we make about their interactions are correct. It is still early days and I am still discovering what the scope is for using this tool. I am looking at what metrics I may be able to use to monitor use and on which I can report, as well as what potential there may be for recognising simple, easily implementable enhancements and improvements to the experience of search at Leeds.

Set-up: systems

At the University of Leeds we use Sierra from Innovative Interfaces as our Library Management System and we have Summon from ProQuest as our resource discovery layer. The Library has been with Innovative for about twenty years, from the early days of GEAC, through Innopac and Millennium to Sierra. The Library transferred the discovery layer from Innovative's Encore to ProQuest's Summon a couple of years ago for the improved central index.

We provide access to both the discovery layer and traditional catalogue. Summon is branded as Search@Library and is the default search box option; Sierra provides 'Classic Catalogue' and this is located on a separate tab. The University of Leeds has an in-house reading list system and a lot of the items on the lists link directly to a bibliographic record in Sierra. Summon and Sierra have been made to work together by the Library Systems Team but the set-up is still not ideal as, although we do nightly updates to keep records up to date, we don't have real time availability information for print stock in Search@Library.

Set-up: teams (IT/Metadata/eResources)

I have mentioned the Library Systems Team and we are incredibly fortunate to have nine people in the Systems Team and within that we have specialists for each piece of software including the repository, reading list system, archives and special collections as well as Sierra and Summon. My main contact has been brilliant to work with. He instigated a series of monthly meetings where he comes and sits in the office with the team and talks through problems people are having; through doing this he has learnt a lot more about the world of cataloguing and we are better able to communicate as he can now speak some 'cataloguer' and we can speak some 'systems'.

The Metadata Team currently have very traditional cataloguing roles. The majority of the work is still monograph-type item cataloguing and journal issues but alongside this is work on special collections and, increasingly, eBooks. One of the big areas still with the Metadata Team is classification. We have a not-entirely-unique system as University College London are also lucky enough to employ the Garside classification scheme. This is the legacy of Kenneth Garside who was Assistant Librarian at the University of Leeds just prior to the Second World War and Deputy Librarian at UCL after the war. Due to the classification scheme, which requires regular updates, and other structural changes we are not currently in a position to go shelf-ready with any of our book suppliers.

EResources are looked after by the Resource Acquisitions Team with the access work being undertaken by a part time Service & Support Coordinator and a part time Subscriptions Assistant. They look after the knowledgebase and ensure that the correct resources are activated in order to feed through into the search systems. We do also have some MARC records for databases in Sierra so that these are discoverable in a variety of ways.

Drivers

The new library strategy '*Powering knowledge and opportunity, 2016-2021*' got underway in September 2016 and one of the actions is on discovery. The headline title of the action is "Review the options for improving the 'discovery layer' to the Library's resources"; within that are sub-actions, the first and perhaps most key is to "review how Summon operates and recommend changes to improve discoverability of content."

My starting point has been to look at what evidence has already been gathered about Library search and assess what is already known about how people interact with Search@Library. The main source of information has been a survey run by the Library Website Project Team where the majority of the free text comments were related to search rather than the website itself.

A project is currently being undertaken to update and improve the Library website. As part of this the project team undertook a month long survey to gather opinions from both students and staff at the University on the website. One of the main findings from the survey was that searching the library's holdings is by far the most important and most frequent task undertaken by all types of visitor.

The report written by the project team has been broken down by responses from different user groups: academic and support staff, research postgraduates, taught postgraduates, and undergraduates. The free text comments provided in the report have been a very useful starting point for me to start looking at possible approaches to the review of Summon. The other element that has emerged from the website project is that they have used Google Analytics to track the usage of the site and draw conclusions about what is useful content and what is never used.

Using Google Analytics

Inspired by the work done by the website project I approached the Systems Team in January this year and asked whether there was any way to track how people were using Summon. I was particularly interested in what search terms were used, which facets were most used (or unused), length of time spent on a search, were any revisions made to those initial searches and did people find what they wanted.

I was given a login for the Summon help centre but the statistical reporting was much more focused on the eResources rather than on the performance and use of the discovery layer itself. It can provide a list of search terms but the information wasn't really at the level I wanted. I was then given access to Google Analytics, which had been used to great effect in the mapping work on the library website project. I was given access to the sites for both Search@Library and the classic catalogue and, at my request, the search terms being entered by users started being recorded for the first time.

As with the Summon help centre, Google Analytics is still a fairly blunt instrument because data protection and authentication methods mean there is only a certain amount of information you can track. The search terms are definitely one of the best things about it for me as one of the first things I've found is that users are pretty good at knowing what to search for and in which catalogue and you have to go a long way down the list to find the first copied and pasted reference. However, one of my key questions, and what Google Analytics can't tell me, is whether the user has found the correct resource. If they revise their search and use new terms then it simply appears in the list as a new search rather than being shown as the same session and a revised search.

A superficial glance at the numbers on the analytics home screen would tell you that the classic catalogue gets nearly twice as many visits as Search@Library. However, this only tells part of the story. As previously mentioned, our reading list system sends users through to the bibliographic record of books in the classic catalogue which accounts for search terms referring to a bibliographic record number appearing in the list of search terms. Also, if a user searches in Summon but finds a print resource and want more details they will be taken through to the classic catalogue.

The other unknown element I'm facing is whether a high 'bounce rate' is a good thing or not. 'Bounce rate' is when someone arrives at the page and then almost immediately leaves again and I would like to know whether it's because they have found the resource they're looking for and have immediately followed a link to leave Summon, or whether they've left because they haven't immediately found their resource. Given the high number of searches for databases that would result in those resources appearing at the top of the list I am currently leaning towards a high bounce rate not being too worrisome but it is definitely an area for further investigation.

There has already been one successful outcome of using Google Analytics: I have fed back to the website team with the first two month's search terms as the key finding is that the top ten search terms in both the classic catalogue and Search@Library are for the same databases. It is now likely that the new website design will incorporate a list of the top ten most searched for databases on the home page to allow quick navigation to these resources. We will keep the search terms under review to see if the easier access to these resources reduces the number of times that they appear in the list of search terms.

Google Analytics can also tell you all sorts of fun but not necessarily useful things such as where your users are located. Unsurprisingly, of the nearly 150,000 Summon sessions in the last month 84% of them were from within Leeds. Perhaps more surprisingly, given that Leeds is a Microsoft institution and 55% of the sessions are from machines running Windows, Internet Explorer is only third on the list of the most used browsers with Chrome and Safari making up 89% of the sessions between them.

Metrics and future development

The easily accessible and fun to report on information of search terms, number of searches, location of users and what devices are being used are useful up to a point. However, in terms of developing a better understanding of user behaviour within the system, ideally I would be able to find out more about how a user develops a search, whether with facets or by revising the search terms they use. I would also like to find a way of assessing whether a high 'bounce rate' is a good thing or a bad thing! I am currently very much in the "fun fact!" realm and I think these elements will be useful for reporting to the wider library and it may be that what I would like to do isn't possible without a more sophisticated analytics tool. However, it is gratifying that the information we are now gathering on the search terms has been useful and has already contributed to the design of the new website.

I have contacted some other libraries to find out what, if anything, they are recording in relation to search and interaction with their discovery layers. There have been studies of user experience relating specifically to resource discovery layers but the methodologies have primarily been very in depth using focus groups or surveys to look at how users approach certain tasks rather than broader brush tracking of common interactions. All the responses I received were from Primo libraries so I currently have no comparisons from Summon libraries but I am keen to make contact with anyone looking at whether the systems can provide insight regarding user interaction with them. It seems clear from the responses I received that this is an emerging area with quite a lot of scope for development; particularly for understanding what users do next, where they go and what we might then extrapolate in terms of the resources we provide, the quality of our metadata and how this can focus our future activities.

e.c.cope@leeds.ac.uk
@ellycope



The National Portrait Gallery's Heinz Archive and Library has a fantastic collection of books about British and international photographers. This brief update explains an ongoing project to catalogue the material in this collection.

Having trained at the Courtauld Institute Book Library, in May 2016 while finishing my Library and Information Studies master's degree at University College London (UCL), I started working part time as Library Cataloguer at the National Portrait Gallery, focussing on the retrospective cataloguing of books on photographers.

As the main centre for research into British portraiture, the Heinz Archive and Library holds more than 40,000 monographs and exhibition catalogues, 150 periodicals, as well as access to electronic resources. Materials include biographies of sitters and artists, and history of art and photography resources. The collection of books on photographers consists of approximately 1,700 biographies, other monographs and exhibition catalogues, covering hundreds of historic and contemporary photographers, from James Abbe to Bettina von Zwehl. More than sixty publications are held on Cecil Beaton, including his diaries. Beaton's influential portrait photography is very well represented at the Gallery. There is also a substantial collection of books on the artist and photographer, Man Ray.

The Archive and Library are used by gallery staff and curators researching exhibitions and collections and increasingly by members of the public who are welcome to use it free of charge by appointment.

The important work of making the Library's collections discoverable online and accessible to the public has been in progress since 2007, with the online catalogue launched in November 2009. A card catalogue is still in use although approximately 70% of the Library's collections have been added to the online catalogue. When the EOS.Web library management system was first implemented, Library staff were able to import very simple bibliographical records for materials acquired since 1994; many of these records still require enhancement. The significant retrospective and re-cataloguing projects carried out so far have been enabled through the generosity of external funders. Increasing accessibility to our collections through good quality cataloguing strengthens the sector's network of support for research and other professional work and hopefully creates further interest in and use of the Archive and Library's resources.

The original funding to catalogue the books on photographers lasted until February 2017, however an extension of the project has been made possible through the generous financial assistance of a private trust. I am therefore continuing to work my way through the collection alphabetically by photographers' surnames, creating or editing around 12 records per day. Detailed records are now online for approximately half of the material, as I have reached 'L'. I am adding RDA 3XX fields to existing or downloaded AACR2 records, updating authorities where necessary and adding detail through subject headings, and contents and exhibitions notes. The work also involves accessioning and processing some of the books, as well as assigning classmarks and labelling them, and identifying items in need of conservation treatment.

The project is great experience for me, using both my art history knowledge gained in my first degree, and the cataloguing skills I developed at the Courtauld and in the Cataloguing and Classification module at UCL. There is a lot of eye-catching material in the collection, with diverse sitters represented in the images. I particularly enjoyed cataloguing the books on Victorian photographers such as Julia Margaret Cameron and Lewis Carroll. It can be a challenge not to be distracted by the amazing images!

The process of cataloguing means that the books are more visible. When the curatorial staff of the Gallery notice a particular book being catalogued, this sometimes gives rise to discussion about the photographers featured. Some of the books have been in the library for many years and it is rewarding to be improving the accessibility of this inspiring collection of material on photographers for more people to discover.

To access the online catalogue, follow links under the 'Research' section of the gallery's website <http://www.npg.org.uk>, or go to <http://librarycatalogue.npg.org.uk/>.

Newcastle chronicle

To paraphrase John Miles, cataloguing was my first love, and it will be my last. I learned the craft of cataloguing from George Ibbs at Newcastle upon Tyne Polytechnic School of Librarianship and Information Studies back in the mid-1980s. AACR2, published a few years earlier in 1978, was our bible, and I have a clear memory of Mr Ibbs intoning “*1.1B1. Transcribe the title proper exactly as to wording, order, and spelling, but not necessarily as to punctuation and capitalization.*” Interestingly, I have a distinct memory of Mr Ibbs speculating on the future computerisation of cataloguing and how this would free up the time of cataloguing staff for other tasks – or simply for redundancy. The use of computers in libraries dates back to the mid 1960s, but over thirty years ago in Newcastle Poly the only computer with a VDU in the department was a BBC microcomputer; our “data processing” classes involved the use of computers which looked like electric typewriters; and Windows, the World Wide Web, the Internet, and Google were in the future.

Streets of London

After completing the Postgraduate Diploma course at Newcastle, I returned to London in 1986 to take up the first of a series of cataloguing or cataloguing-related posts in libraries including the BLPES (British Library of Political and Economic Science at the London School of Economics), the British Library (before the move of its cataloguing operations from Sheraton Street in Soho to Boston Spa), and the British Architectural Library at the RIBA (Royal Institute of British Architects). At BLPES, we made use of the MARC file on BLAISE (the British Library Automated Information Service) to copy and customise UKMARC records for the local catalogue. By the time I was working as a cataloguer at the British Architectural Library in the early 1990s I was cataloguing directly onto the character-based STATUS information retrieval system, but this was being replaced by Sirsi’s Unicorn library management system around the time I left the RIBA at the start of 1996.

From the Smoke to the dreaming spires

Almost ten years after graduating from Newcastle, I moved from London to Oxford to take up a fixed-term academic-related post in the Cataloguing Division of the Bodleian Library, and I have worked in Oxford ever since. The position I held had been created to assist with the migration of OLIS (the Oxford Libraries Information System) from IBM’s DOBIS-LIBIS to the Geac Advance library management system. I was part of a small team that tested the cataloguing software and assisted with its configuration, wrote documentation, and provided training for cataloguers in University, college and departmental libraries within the Oxford University system. Geac was one of the first system vendors to offer a Windows-based Z39.50-compliant GUI cataloguing module – GeoCat – and the Head of Catalogue Support in Oxford worked closely with the developers at Geac HQ in Canada to ensure that by the time it was implemented in Oxford on 2 April 1997 GeoCat was one of the most sophisticated and user-friendly cataloguing modules on the market. Oxford had already taken the sensible decision to adopt AACR2 and USMARC when automation began in 1986¹, so the migration of cataloguing data was comparatively straightforward. When the Geac system was implemented, we took the radical decision to load the entire Library of Congress Names and Subjects authority files into the local system, and worked closely with Geac developers to enhance the Advance authority loader to cope with this innovative albeit labour-intensive approach to in-house authority control².

¹ Burnett, Peter P. (2000) Emerging from the bibliographic wilderness: catalogue automation in the Bodleian Library, University of Oxford. *Cataloging & classification quarterly* 30(1): 51-72.

² More details of the migration to the Geac system can be found in: Burnett (2000, 60-65).

The OLIS database was created in 1988. (The public interface to this database was renamed SOLO – Search Oxford Libraries Online – when the Primo resource discovery layer from Ex Libris was implemented in 2008, but the OLIS acronym still survives on the staff side in the online OLIS Support Centre and the OLIS News blog.) Oxford University was by no means unique in going down the path of co-operative cataloguing into a union catalogue, but I believe that OLIS cataloguers receive a level of support which helps to ensure that the quality and consistency of the OLIS database is maintained. These cataloguers range from solo librarians in the smaller colleges to the 25-odd staff in the Bodleian Libraries' Resource Description section. The support they receive takes the form of initial training and documentation, refresher training on specific topics, and ongoing support by email or very occasionally over the 'phone. The ongoing support is particularly important for librarians in the many smaller libraries, for whom cataloguing is just one relatively minor aspect of their work.

The current population of Oxford cataloguers is fortunate to have excellent support from two dedicated members of Bodleian Libraries staff. General cataloguing support is provided by the Catalogue Support Librarian, and technical support by the Systems Support Librarian, who has exploited the configurability of the Aleph cataloguing module (implemented in summer 2011) to make the day-to-day work of OLIS cataloguers as easy as possible and to ameliorate the migration in 2013 from AACR2 to RDA³. In addition to the two staff mentioned above, the Bibliographic Maintenance and Authority Control (BMAC) team of four staff provides support for OLIS cataloguers in around 100 different libraries. The BMAC team also manages the authority loading and performs database-wide catalogue maintenance and authority control tasks. Of course, any system that numbers around 200 staff with the ability to input records into the shared database is going to face challenges of quality control, but these are mitigated as far as possible by the work of the support staff and BMAC.

Managing current cataloguing in the Bodleian

In November 2010 I was appointed to the position of Head of Resource Description (aka cataloguing) in the Collections and Resource Description (C&RD) department of the Bodleian Libraries. This involved managing around 25 staff in four teams involved in cataloguing the legal deposit intake, purchased books (English and European), serials, ebooks, ejournals and other resources, plus authority control, catalogue maintenance, and cataloguing documentation and training. I also held the *ex officio* roles of Chair of CatSIG, the OLIS Cataloguing Special Interest Group, and of the OLIS Cataloguers' Forum. The former brings together a small group of representatives from the different types of library and different subject areas in the Oxford University system and meets twice a term (once virtually) to discuss local cataloguing policies and review developments in the wider world of cataloguing. The latter is a forum for all staff that catalogue into the OLIS database and meets termly to learn about the latest developments in cataloguing – local, national, and international – and to ask questions and discuss their impact on cataloguing in Oxford.

I resigned from this post in February 2016 and left the Bodleian in May. At the time of writing, my understanding is that attempts to fill the post were not successful, and as a result the post of Head of Cataloguing in the Bodleian Libraries has been abolished as part of a mini-restructuring of technical services operations. I believe this means that for the first time since the early 1930s, when Strickland Gibson (1877-1958) fulfilled the role of "Superintendent of the catalogue of printed books" (among other duties)⁴, the Bodleian does not have a dedicated Head of Cataloguing. Does this surprising development tell us anything about the future of cataloguing? Anecdotal evidence suggests that other libraries are also experiencing difficulties in filling cataloguing posts at all levels. Or is it simply a reflection of the very high cost of living in Oxford?

³ More details of the implementation of RDA in Oxford University can be found in: O'Reilly, Bernadette (2013) RDA at Oxford University. *Catalogue and index*, 173 (December 2013): 50-56.

⁴ Information gleaned from the Bodleian Library's *Staff manual* of 1933, with thanks to Robert Bruce for assistance with this research.

It's cataloguing, Jim, but not as we know it

The nature of the work of the English cataloguers in the Resource Description section is changing. Over the past few years there has been a noticeable shift away from creating records with book-in-hand to the checking of reports of batch-loaded records from the BNB database. The implementation of Non-Print Legal Deposit following the passing of the *Legal Deposit Libraries (non-print works) regulations* in 2013 has led to a reduction in the volume of printed monographs deposited with the Bodleian and the other Legal Deposit Libraries, and a corresponding reduction in the number of items requiring cataloguing. On the other hand, the increase in research outputs and research data managed for the University by the Bodleian Libraries in its institutional repository (ORA, the Oxford University Research Archive) means that staff with traditional skills are being encouraged to consider new roles, as indicated in a recent job description for a Metadata Assistant based in the Bodleian Digital Library Systems and Services (BDLSS) department. This post included among its responsibilities “*enhancing the metadata for [research outputs] and checking names to ensure that material is associated with the correct authors where more than one author shares the same name.*” The advert continued “*This post provides a great opportunity to transfer the ‘traditional’ bibliographic skills of cataloguing and authority control to the newer and still-developing area of research output data management.*” There have been some concerted efforts over the past few years by managers in C&RD and BDLSS to work more closely together to share expertise and staff resources for metadata work, and a handful of staff have moved from posts in one department to another or have worked on projects that extend their traditional cataloguing skills into using other metadata standards including Dublin Core and MODS. These sensible initiatives must surely continue, and cataloguers who currently work with MARC and AACR2 or RDA may future-proof their positions and improve their employability by seeking opportunities to gain knowledge and experience of newer and emerging metadata standards and systems.

On the topic of new standards, I don't think that anyone can predict at the present time when (or if) BIBFRAME will replace the MARC formats. A recent thought-provoking paper by an experienced cataloguer in the US contends that BIBFRAME will not be widely adopted⁵. I confess that I have not followed closely its development, and cannot comment on its merits or demerits. My only observation, speaking as a former cataloguing manager, is that its implementation is likely to be driven by the systems vendors. If and when BIBFRAME is integrated into next-generation resource management and discovery systems such as Ex Libris's Alma, cataloguers may find themselves with no option but to use it when the time comes to migrate to a new system.

Taking refuge in the past

One future for those who love cataloguing lies in the creation of records for rare or antiquarian books and other printed materials. Rare books cataloguing has in some libraries been regarded as the poor relation of modern cataloguing, with fewer resources being expended by libraries. After leaving the Bodleian last year, I was fortunate to find work in Trinity College, Oxford, cataloguing rare books. Each week I discover items for which no record exists in the OLIS database, in spite of a number of initiatives since the late 1980s to create online catalogue records for early printed books in Oxford college and departmental libraries, and in the Bodleian and faculty libraries. Extending the search to the RLUK (Research Libraries UK) database, the ESTC (English Short Title Catalogue) database, the HPB (Heritage of the Printed Book) database from CERL (Consortium of European Research Libraries), or OCLC's WorldCat does not always yield a result – or if a record is found, it will usually require significant enhancement to bring it in line with the current standard for rare books cataloguing in Oxford.

⁵ Edmunds, Jeff (2017) “BIBFRAME as empty vessel.” <https://drive.google.com/file/d/0B1IKJYVwLwHyX1VnbIJFZ3EtS1U/view> (visited 5 March 2017).

Copy-specific cataloguing, whereby the rare books cataloguer describes those aspects of a printed book which are unique to the copy in hand, is an area of rare books cataloguing where there is still much work to be done. Copy-specific information can include binding, bookplates, provenance, imperfections, manuscript additions and insertions, hand-coloured illustrations, previous shelfmarks, whether the item is bound with other distinct items, etc. Rare books cataloguers have been recording this information in the OLIS database for nearly thirty years, but there is still much work to be done to expose the richness of hidden collections for the benefit of scholars and other researchers. I find rare books cataloguing a very satisfying use of my skills, and feel very grateful to be able to focus on this specialised but important area of work when the future of cataloguing in general is uncertain. Ironically, the future of cataloguing – for this cataloguer at least – lies in the past.

Where to start in understanding metadata is an often-discussed topic for both the designers of library school curricula and practitioners who want to expand their knowledge from everyday cataloguing to the wider theoretical concepts in which their practice sits. Richard Gartner's book provides a fine introduction, whatever the reader's own starting point.

One of the strengths of the text is its brevity. At only 114 pages, it seems difficult to imagine exactly how it will, as it's back-cover blurb claims, *"take us on a journey through the centuries-old history of metadata up to the modern world of crowdsourcing and Google, showing how metadata works and what it is made of."* Yet Gartner manages it, drawing on his longstanding experience as a Digital Librarian and more recent work as an academic at King's College London and, as part of his remit at the Warburg, at the newly-established Digital Humanities centre at the University of London, DH@SAS.

A lot of the concision is facilitated by the use of (mostly colour) illustrations, diagrams and catalogue records, which are nicely done. The historical emphasis, although important in the first two chapters, is kept to the minimum required for us to understand what is happening now, in the 21st century, and I am sure this approach will appeal to those who are not interested in history *per se*, as well as being helpful for those of us who work in historical disciplines. Do not be deceived or put off by the titles of chapters 2 and 3 – 'Clay Goats and Trees: Metadata Before the Byte' and 'Metadata Becomes Digital.' The author is not telling us about our past merely because it is interesting, but as part of a narrative that leads us through a philosophy of metadata towards an ideological understanding of what it is, and why our mission as its creators and curators is vital to the creation and understanding of our very culture.

All of the usual elements of a book on metadata are present and correct – card catalogues, MARC, taxonomies, ontologies and networks – but Gartner is writing with a message that makes his work interesting, challenging and worthy of wider discussion within the cataloguing community. Metadata is not neutral – we are making ethical and political decisions all the time. As Gartner puts it in his final chapter, titled appropriately 'Knowledge and Uncertainty,' *"Metadata has an essential but often invisible role in the way we build our knowledge. It allows us to bring together its atoms, the single units of data into which we can divide the world, into the information that gives them meaning. It then lets us assemble this information into knowledge."* (p. 107). Further, *"We can find patterns of meaning to help us fulfil the very human need to interpret what we have and make sense of it. Metadata lets us forge these links and construct our edifices of knowledge from the smallest bricks of data."* (p. 108).

You don't have to agree with the author's opinions, but I would urge you to read this book if you are new to cataloguing or if you are old in cataloguing and world-weary about our professional mission. It's a book to provoke your own thoughts, and quite possibly to give to your manager if you suspect they are unsure why metadata, why cataloguing and, therefore, why you and your team matter.



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

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

Submissions: In the first instance, please contact the Co-editors:

Karen Pierce: PierceKF@Cardiff.ac.uk

Deborah Lee: Deborah.Lee@courtauld.ac.uk

Book reviews: Please contact the Book Reviews Editor:

Anne Welsh: a.welsh@ucl.ac.uk

ISSN 2399-9667

CIG website: <http://www.cilip.org.uk/cataloguing-and-indexing-group/catalogue-index>

CIG blog: <http://catandindexgroup.wordpress.com/>

CIG twitter: @CILIPCIG

Tags from the CIG blog on specific areas of interest:

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