Welcome to this bumper issue of C&I which features a selection of papers from this year’s conference held at Swansea University 31-August – 2 September 2016. Having attended quite a number of CIG conferences (the first being Newcastle in 2002) I feel able to comment on the quality of this year’s event. The papers were extremely relevant to delegates and changes to the format of the event created choice, variety and interaction. James Baker (Library Carpentry) gave an excellent keynote speech and we also enjoyed a conference after dinner talk by book artist, Sam Winston.

The theme was innovation and discovery. We have papers from practicing cataloguers which look at workflows and innovations in specific areas of cataloguing: Clifford, Quayle, Dickinson, WHELF, La Spada. We also have papers which are more focused on metadata and discovery: Danskin, Wright, O’Reilly, D’Ambrosia and Staniforth. The importance of good metadata for discovery is a consistent message.

There is also a report on the Equality and Diversity of the CIG member network which was initiated by the CILIP leadership project.

I have now served my “time” on the CIG committee and therefore this is my final issue as a co-editor. I began as editor in January 2013 (C&I170) and have co-edited 16 issues of the journal. Whilst it has taken up some of my free time, it has also been one of the best professional activities of my career and I am really pleased that it is now an Open Access journal. The job of a C&I editor is made a little easier when there is a flurry of offers after the call for papers, so please contact your editors, Karen (and whoever takes my place), with ideas for articles in the future.

Best wishes
Helen

Helen Garner (h.j.garner@shu.ac.uk)

Karen F. Pierce (PierceKF@cardiff.ac.uk)
All too often we cataloguers can get caught up in the details of cataloguing, cataloguing in certain ways because of the standard we use, and forget the real point behind why we catalogue in the first place – for our users benefit, to make our material discoverable. We at the University of Kent decided it was time to take a step back and look at our catalogue as if we were users, consider what we would need from our catalogue and our metadata, and how could we as cataloguers achieve this? We identified our key users. Our most obvious group is our students. Undergraduates want to be able to find something useful for their assignments quickly and easily, where postgraduates usually want more detailed information about the item as their research areas are more niche. Staff need enough info to be able to answer queries – at the moment with the variations in historical cataloguing this can be difficult. Outside academics and researchers will want as much information as you can give them to save a wasted and potentially expensive trip if the item is not what they wanted – particularly foreign researchers. We have had researchers visit us from the universities of Niagara, South Florida and Texas Christian, which has been possible partly because of our metadata. Our partner institutions want a good quality of record to display on our resource discovery system. One benefit we bring to our partnership with Rochester Cathedral is that we provide high quality metadata to increase their discoverability and their presence globally.

The University of Kent moved to cataloguing in RDA from AACR2 in 2013. This was a macro process where the RDA carrier fields and the 264 field were inserted into all our records, and some abbreviations were expanded. On site work then continued in RDA. There are some very clear immediate benefits to cataloguing in RDA to enhance discoverability. The 264 allows for clearer labelling of the book production process. The various 7XX fields are now limitless – previously an author of a book section, or former owner of an item may not have been included in the metadata, now they can be. Removal of Latin terms and expansions of abbreviations are extremely valuable to our students, most of whom will not know what they refer to. Terms such as ‘polyglot’ are specialist, which undergraduates, and to some extent postgraduates, are not. This is our specialism, and we should be using our knowledge to make metadata clearer for them.

Special Collections items have a particular set of discoverability issues that is not just based around changing standards. Historically, Special Collections cataloguing has been hugely varied and inconsistent, partially because we have had no previous rare book cataloguing specialism. Many of our records have been transferred between more than one library management system, which has caused information to be lost or moved around fields, making them messy to look at and remarkably unhelpful. It can be highly frustrating to know that something exists in your collections, but cannot be found on the catalogue, or a record you once found, but have no idea how you managed it, is now nowhere to be seen. Our staff have found it nigh impossible to run any kind of report on Special Collections books, as there is no consistent field from which we can pull information. Not only have things been done differently historically, sometimes it has been done inaccurately as well.
We wanted to create a policy for Special Collections to ensure that future and retrospective cataloguing could focus on the same information and become consistent, but we held off, as RDA has yet to develop a rare books standard.

Our hand was forced in making a decision when we were approached to assist in a cataloguing project on behalf of Rochester Cathedral, largely as a result of our partnership work with Canterbury Cathedral. In early 2013, the Heritage Lottery fund awarded Rochester Cathedral a grant of £3.55 million to make its ‘hidden collections’ more accessible to the public. Part of this included the addition of disabled access throughout the cathedral, the refurbishment of the crypt and library, and the cataloguing of their books. The library is made up of around 7000 items, including roughly 2500 dating from before 1901, and has never been catalogued to currently recognised standards. As part of the project required us to catalogue the pre-1901 items using MARC21 and RDA, we felt the urgency of deciding how precisely we would achieve these aims, and knew we would need to develop a policy.

Our initial concerns surrounding this included the amount of research we would need to do in order to create a policy that encompassed all we wanted to achieve, and that would be transferable to our own collections as well. We needed to work in RDA, but wanted to use elements of DCRM(b) to maximise the information our users could see, after all, we do not know what they are looking for. We decided a group approach was the best way forward, with the senior library assistants of Special Collections and Metadata, and myself and Josie Caplehorne, who catalogue Special Collections items here at Kent. We also consulted Karen Brayshaw, the librarian of Canterbury Cathedral, to tap into the unique well of knowledge we had available to us. The three cataloguers decided to create the ultimate bibliographic template, encompassing every field we could ever need, with guidance and examples of how to use them appropriately. Whilst useful for working out how much detail we needed to go into for Rochester, it also meant we would have an excellent resource for future staff training. We divided the fields roughly into thirds, each taking a section to investigate thoroughly. Results were three separate sets of guidelines that were then amalgamated into one super guide.

We initially tested our work during the cataloguing of our own David Lloyd George collection, which had previously been partially, but poorly catalogued. The results were fantastic – the number of requests to view Lloyd George items in our Reading Room increased dramatically the more we catalogued the items.

The Holy Bible, 1717, ‘Vinegar Bible’, Rochester Cathedral Library
Whilst pleased with our work, we didn’t want to rest on our laurels, and invited members of the cataloguing team at Lambeth Palace down to Kent to talk to us about this style of cataloguing. In the knowledge we were shortly to catalogue a significant ecclesiastical collection, it was hugely beneficial to have experienced cataloguers in this area talking to us about what they do. We showed them examples of old catalogue records and Lloyd George records for comparison, to show them the changes we had made, and to get their opinions on anything else we could do to improve our work. Their visit resulted in a few tweaks to our policy, and when we were convinced it was a good as it could get, it was applied to the Rochester project. As we had taken the time to test the policy on Lloyd George, we had no significant issues with it when cataloguing the Rochester collection.

It can be hard as cataloguers to establish how your work is benefitting your users. Here at Kent, members of the metadata team have had the opportunity to work with materials at outreach events, such as exhibitions and seminars, to talk to people about the material – after all, who will know a collection better than the person who has catalogued it in great detail? As part of the publicity work for Rochester, several exhibitions have been held at the Canterbury and Medway campuses of the university.

Here we had the best opportunity of talking to the people who mattered – our users. This enabled us to ask and answer questions about the collection, meaning we could establish precisely what it is our users need from our metadata. One of the most interesting outreach opportunities took place in the middle of June. As part of the Medieval and Early Modern Studies Summer Festival, Josie, the most experienced cataloguer on the project, ran a sort of workshop, talking to people about the role of a cataloguer, explaining precisely what information we look for and why we include it. People are surprisingly interested in this work. If we can explain to our users how we work, it will give them a better understanding of how they can use our catalogue, and what will give them the most useful search results. A visitor from the Czech Republic, Dr Lenka Panušková, was particularly happy to learn we were cataloguing provenance to the greatest detail we could manage, and this is a common area of interest. Previously, the quality of Special Collections records meant we often received enquiries asking if our copy of a work had any annotations on it, but now the users will be able to get this directly from the catalogue records.

Seminars run for our students using Special Collections material is another fantastic time for our users to discover what we have. Historically this has sometimes lead to frustration, as items they saw in our sessions were next to impossible to discover on our catalogue. This has, slowly but surely, been improving as we work to improve our metadata.
The most positive outcome of any kind of outreach is when you get your users interested and enthusiastic to help you. One of our MEMS students, (see below), has written a detailed work on the St Augustine manuscript from the Rochester collection, which will be published in the future. Professor James Carley, an expert in early bibliography, has studied the Rochester collection in great detail, gave a talk at Kent on the history of the library, and has published a paper on the subject. Dr Panuškova and another Kent MA student have become interested in working Kent’s mysterious astrological tract, an undated manuscript in Latin with astrological charts which none of us know much about. MEMS students have also been instrumental in transcribing manuscripts from our local history collection. This is hugely beneficial from a metadata perspective. As cataloguers, we specialise in what we do, but sometimes our users don’t understand that our area of expertise is our collections, and not the subject area or time period. The more we can work with experts the more detail, and most importantly the more accurate detail, we can add to our metadata. For unique works like those in the Rochester collection, this is incredibly significant, as we will not necessarily even know if we are missing important information. It can also potentially save our particularly rare, valuable or delicate items from being over handled, if a researcher can get all they need from our records without looking at the item.

Kent has found great benefit in the use of RDA, and the creation of our own hybrid of RDA and DCRM(b), in increasing the discoverability of our collections. Records are easier to find and contain more information than they ever have before. They are also much more understandable for those who are not book specialist, which is most of our user group. Our work with Rochester Cathedral has illustrated the important of detailed cataloguing, ensuring that our users can truly understand the significance of our material. Giving our users an understanding of how we work and what we look to include in our metadata can improve their searching skills. Finally, and it cannot be stressed enough how important we at Kent have found this, talking to and involving your users, utilising their knowledge and expertise, can be a huge benefit to your work. We can’t know everything. Everyone has their own area of expertise, and making use of those who have specialist understanding means we can include the most accurate and appropriate detail in our records, which in turn makes our material more globally discoverable.
Metadata on the move

Off to Paris next week, but whoops! I forgot to renew my passport. No problem – I'll borrow my son's. Of course, he is a young man, fair-haired and six feet tall, while I am female, darkish, shortish and oldish, so I'll have to make a few adjustments to the photo and to any other elements which relate to my personal description.

What could possibly go wrong?

Well, I might end up in gaol, which would prevent me from travelling at all. If the Border Agency happens to be cutting corners that day, I might avoid detection and get to France. But that could cause another set of problems. Records might then indicate that my son is in France, while I am still in the UK. If someone really needed to get hold of one of us, they might get the wrong person, or no one at all.

Because, even if I change the descriptive bits on the passport, it still has a serial number which is associated uniquely with my son, and so goes with his description, his attributes, his origins and travel history, not mine.

Nowadays catalogue records are travellers too. Many agencies continually import records from WorldCat, RLUK, the Library of Congress, the British Library and others, and export their own records back to WorldCat and RLUK – to say nothing of harvests and conversions from non-MARC metadata. And records, like human travellers, have identities and histories, which are enshrined in various serial numbers and other identifiers. These identifiers associate each record with a particular resource, and by implication with key aspects of its description, for instance its title, content, date of issue and physical format. Together the identifiers and descriptive elements act rather like a passport. The identifiers should make it possible for agencies to export their data without fear that it will be wrongly associated with data for other resources and to import data without fear that it will be of the wrong type or for the wrong resources.

For instance, the record below (p. 7), from OCLC’s WorldCat, has the following identifiers (in red):

- OCLC system number (035); this will be preserved if the record is downloaded to another database
- National bibliography number (016); other records might have a BNB number (015) and/or a Library of Congress control number (010), doing similar jobs.
- Codes of the agencies which created or edited the record (040 $a, $c, $d), providing it with a sort of individual origins-and-travel history.

Just as for human travellers, if the descriptive elements within a record change radically they conflict with that record’s identifiers, with the result that the record either fails to travel at all or is associated with the wrong resource and misleads people about where that resource may be found.

Nowadays records even have something like nationalities, the language-of-cataloguing-agency recorded in 040 $b, in this case French (in green). This is not the language of the resource, but the language used in the non-transcribed elements of the record, e.g. for physical description and notes. WorldCat now has records from many different language communities, and, to allow users to find records in their own language, has ruled that a change to the language-of-cataloguing requires a new record. That means that the identifying numbers in a record’s 035 fields are associated not just with the resource catalogued but also with the language in which it is catalogued. National bibliography numbers may also imply a particular language of cataloguing.
Has our cataloguing culture kept up?

Although I am employed by the Bodleian Libraries in Oxford, I work with all the cataloguers in Oxford University’s Libraries Information System (OLIS), providing training, documentation, advice and a bit of quality control. Altogether I work with nearly 200 cataloguers in nearly 100 libraries scattered over several miles, half of which (college libraries and some departmental libraries) are completely independent of the Bodleian. I can’t be looking over everyone’s shoulders, so for quality control I have to depend on reports, when I know what to look out for, and otherwise on serendipity – just happening to come across records which indicate misunderstandings or knowledge gaps. So sometimes I get surprises.

All the big surprises in the last few years, the drop-everything-to-fight-this-fire surprises, have involved the kind of data which is used for managing records both internally and externally – the elements which control things like whether a record should be edited, or upgraded, or overwritten, or exported to COPAC or WorldCat and, particularly, how our exported records will be handled by the receiving agency. Fires of this kind have proved particularly difficult to fight – smouldering unnoticed for years, taking hold in several locations at once and with a trick of breaking out afresh when we hope they have been damped down.

The two biggest surprises related to ‘passport control’, and I would like to share what we learned from them, because they indicated that our cataloguing culture had not kept up with the changing cataloguing environment. They showed that our cataloguers found it much easier to appreciate the importance of good descriptive data and access points than the importance of the kinds of codes and identifiers which are primarily intended to be read by machines for automated processes, especially when the processes involve databases other than our own. But this kind of data is absolutely key to our ability to meet the growing challenges of dwindling resources and the growing opportunities of linked, machine-actionable data. We know that these problems are not confined to the Bodleian. Some of our faulty records had been copied from other databases with the problems already present – from OCLC or RLUK, even a few from the Library of Congress and the British National Bibliography.
I am not at all intending to criticise my own colleagues or cataloguers in other agencies. When I was first trained to catalogue, just over a decade ago, I did not need to know very much beyond AACR2, MARC21, LCSH, our local software, and some fairly basic information about the external databases from which we could copy records. Now our cataloguers are expected to understand and make best use of a vast, varied and ever-changing network of data and processes, both local and external. They have to be up to date with what to trust, what to check, what to adapt, what to avoid entirely, so that they neither waste time on unnecessary checking of good data nor accept data which is not fit for purpose. It’s a big ask. Senior staff may design and document ‘efficient’ new streamlined workflows and algorithms, but that is not enough – the workflows and algorithms will not work efficiently until we win the hearts and minds of all our colleagues to care about all the kinds of data which the workflows and algorithms use.

Records from foreign-language agencies

The first shock hit us in the summer of 2012, just when we were focusing on our preparations for RDA implementation and really, really did not want to have to think about anything else. We noticed a record which did not use Unicode and so displayed oddly in our system. Its 040 $b and $e showed that it was created by a German-language agency to a German standard (Regelwerk für Alphabetische Katalogisierung Wissenschaftliche Bibliotheken), and in other respects it was unsuitable for our use. For instance, the physical description was in German and there were no English-language subject headings.

But that was not the shock.

Initially we thought that the record was more or less a one-off, and that all we needed to do was to remind our cataloguers that OCLC included non-English records, and make sure that everyone knew what to look out for and avoid. The shock was when we were told by several managers that their cataloguers could not get through their workload without downloading non-English records, and had been downloading them for years. In fact, there were already several thousand records in OLIS coded as non-English. We just hadn’t noticed them before, among our millions of records. Most had been more or less fully translated and adapted to English-language standards, but were still coded as French, German, Dutch, Spanish, Italian, while a few had not been translated or adapted at all, and so still had physical descriptions and notes in languages other than English.

Because our attention had been absorbed by RDA and, before that, a system migration, we had not given much thought in the previous few years to the fast-increasing availability of foreign-agency records, and many of our cataloguers were actually unaware of the significance of the code in 040 $b. It had hardly occurred to me or my immediate colleagues that anyone might be tempted to use a record which had funny 300 fields and used unfamiliar standards and authorities and subject headings – surely it would be simpler to catalogue originally? That was a failure on my own part to engage with this new source of potentially valuable data, because when original cataloguing of a foreign-language resource would involve lots of diacritics or nonroman scripts and transliteration, adapting a foreign-language record can be an attractive option.

It is obvious that untranslated or semi-translated records are not fit for purpose. Users might well be perplexed by finding in an English-language catalogue a physical description such as ‘24 cm met CD’ (‘met’ is Dutch for ‘with’) or a note such as ‘Lizenzpflichtig’ (German for ‘Subject to license’).
And it is not just the language itself which might cause problems for users: foreign-language records often use different cataloguing rules, different versions of MARC, different name authorities and different subject systems, so a great deal of editing may be necessary to make sure that users can find the resource in the usual way and discover its relationships to other resources in the collection. But if a record has been thoroughly translated into English and edited to our usual standards, does it really matter if it still has in 040 $b a code which claims that it is in a language other than English and/or has a number in 035 which identifies it with a record in a language other than English?

It would not matter if the record was never going to travel anywhere else ever again; but in fact all our finalised full-level records (and many less than full) are exported to OCLC and RLUK, and so it matters in the following ways:

- OCLC will often reject entirely a record which has an 035 number identifying it with an existing OCLC record but differs from that record in key elements such as language of cataloguing or extent, so changing ‘ger’ to ‘eng’ in 040 $b or changing ‘Seite’ to ‘pages’ in 300 $a without giving the record a new identity might mean that the record does not reach WorldCat, and WorldCat users will not know that we have a copy.
- If translated records with unchanged language codes do slip into OCLC, they will be a snare and a delusion. For instance, a Spanish-language agency which wants to harvest Spanish-language records, selected by language of cataloguing (040$b=spa), might actually get a proportion of records which have been translated into English. This is particularly disturbing because it feels like a breach of trust and a betrayal of our common values and objectives. If agencies cannot be reasonably confident that data which is labelled as being of particular types really is of those types, all our initiatives to maintain and improve services by exchanging and linking data will be undermined.

It’s much the same as if I tried to travel on my son’s passport: either I wouldn’t get to travel at all, or I would slip through the checks and then people who wanted to find my son might get me instead.

We had to get ourselves sorted fast:

- For the thousands of existing OLIS records with non-English 040 $b, we could only do a rough global correction, because there was so much variety as to how far they had been checked and/or translated; so, sadly, all those records had to be downgraded.
- A systems colleague designed a brilliant fix to make it possible to adapt downloaded foreign-language records safely. It recodes the records correctly as new English-language records, removes the history, and adds a local field with a warning that the record must be checked thoroughly for conformity with MARC21, RDA, LCSH, etc. Cataloguers are not supposed to delete this field until the checks are completed, and meanwhile it prevents export.
- I added a new check to our record-checking software (Marc Report) to warn cataloguers if they saved a record as full-level and ready for export while it still had a code other than ‘eng’ in 040 $b.

1. Some delegates at the CIG 16 conference were unhappy about removing the history of work done by other agencies, and it certainly does not feel good to remove an indication of intellectual input; but a new record by definition has no history of its own, and it is not usual to record ‘ancestry’ for anything less than a whole record. A cataloguer who re-purposes some of the data from an existing record takes over responsibility for that data, just as s/he would if s/he created an original record but saved time by copying over data from some other record field-by-field. OLIS does actually keep some local data which indicates the record’s previous incarnation, and Alan Danskin of the British Library mentioned that ancestry could be recorded more formally by using MARC field 038, Record Content Licensor, but this is not common practice. Perhaps this is an area for more discussion among the cataloguing community. The situation has crept up on us without much opportunity to think through all the implications.
But cleaning up the bad data and making safe procedures available was only the beginning. We still needed to make our cataloguers fully aware of the importance of using the fix and making the checks, and that has required a very serious and sustained effort: not just documentation, automated warnings and general reminders, but in many cases person-to-person explanations, talking through how users and staff, both in OLIS and in other agencies, might be misled or inconvenienced, so that the issue no longer seemed like a mere technicality. This ‘humanising’ of record-management data is particularly difficult for us because our cataloguers are so scattered, but it is very worthwhile.

After four years things are much better, but we still occasionally come across records where a cataloguer has made the warnings go away by making little edits to the elements which trigger warning messages, for instance by deleting 040 $b, rather than using the proper fix and procedures to assign a new identity.

‘Unresolved’ records

After our last system migration it was not possible to set up a regular export to OCLC for quite some time. When we eventually exported the backlog, several thousand records were rejected. Many of these were translated records which had retained the 040 $b or 035 of the original, as explained above, while others were technically incorrect, such as records for mathematical or scientific resources with untransliterated Greek characters in the title fields. But very many were rejected simply because they had been derived from OCLC and retained the 035 fields of the master records from which they were derived, but they had been edited in ways that made them seem to be for different resources from those covered by the master records, with the result that OCLC’s deduplication algorithm could not process them.

In a minority of cases the edited records really were for exactly the same resource as the original OCLC record but had been improved out of recognition (e.g. brief OCLC records enhanced to antiquarian standards), and in such cases the only solution is to improve the OCLC record correspondingly. Far more often the records really had been edited to match different resources, for instance by changing the edition statement or ISBN or publication date or carrier or format or by extending a single-part record to cover a multipart resource.

Like the foreign-agency records, these illegitimately adapted records cause two types of problem:
- If our records are rejected, WorldCat users will not discover that we have those resources.
- Some adapted records do get through, cluster with the wrong master records, and give users the impression that we have a resource which we do not have, wasting their time and disappointing them.

As with the foreign-agency records, what was really startling was that records were being illegitimately adapted not just occasionally by oversight, but quite routinely by experienced and careful cataloguers. In some cases a single OCLC record had been downloaded as many as four times for separate editions of a work, with the result that we had four different OLIS records all with the same OCLC identifier in the 035 field.

On the next page is an example, showing key fields in an OCLC record and in an OLIS record which was derived from it and carefully enhanced to antiquarian standards. The OCLC identifier in field 035 is retained in the OLIS record, but the title proper, (lack of) edition statement, place, publisher, date and physical description are all very different, showing that it is being used for a quite different resource. In effect, it is using another record’s passport. When it tried to travel to OCLC it was caught and sent back.
The solution for this problem was in principle quite simple. No new processes or procedures were needed, and the existing documentation described the proper processes and procedures correctly. What was lacking was (i) a clear understanding that copying a whole record, identifiers and all, is not at all the same thing as just copying a lot of useful bits of data, and (ii) an appreciation of the human cost of using the wrong processes—the frustration and wasted effort for users, and the deflection of precious staff resources into laborious record-by-record analysis and repair. This particular hearts-and-minds campaign has been running for less than a year. There has been a great improvement, but some cataloguers do still sometimes forget, particularly generalists who catalogue only occasionally.

These are just two examples of the human cost of paying too little attention to the parts of a record which are intended to facilitate processing by machines. The importance of such elements can only increase with the challenges of dwindling resources and the opportunities of linked, machine-actionable data. If we are serious about sharing and caring, we nowadays have to be prepared to share and care on a scale so large that it can only be managed by mechanised processes. By exchanging data on a really large scale we keep our operations viable, enrich our services and support other information providers.

So cataloguers must learn to care about how they talk to machines. And for that to happen, people in jobs like mine must care about how they talk to cataloguers and remember to present mechanised procedures and processes in the context of the human needs and aspirations which they serve.

---

2. A cataloguer who wants to improve an existing OLIS record with data from a better record has to copy over that data field by field, but that is less tedious than it sounds, because our Aleph system allows multiple adjacent fields to be copied simultaneously. The important thing is just to avoid using the ‘Duplicate Record’ command.
This article presents the experiences of 3 members of the WHELF http://whelf.ac.uk/ Shared LMS Consortium who implemented ExLibris’ Alma & Primo products during 2015/16.

While NLW’s collections cover a broad range of materials, including archives, photographs, film, sound recordings, maps and many more, print materials still form the vast majority of our holdings, particularly due to our status as a legal deposit library. Over 2,000 printed items arrive here each week (121,186 items in total were received for 2015/16). Journals, which comprise under half the total, are received shelf-ready, requiring little intervention other than for new titles. The books however are individually processed by staff, and with such large volumes it is vital that the LMS allows us to implement efficient and reliable workflows.

The majority of books are processed by a team of three following an end-to-end workflow; staff search for a record via Z39.50 (usually from the British Library), and having found a suitable match they import the record, create holdings and item records, apply labels and then shelve the books in batch. This workflow was easily implemented in Alma with very little change involved for the staff.

A proportion of each weekly delivery, those books from Welsh imprints or with titles beginning I-O or W-Z, are filtered out for cataloguing from scratch by more senior staff. The high-quality records that are created for these items are our contribution to the Legal Deposit Libraries Shared Cataloguing Programme to create the British National Bibliography http://www.bl.uk/bibliographic/natbib.html and in 2015/16, we exported 4,228 records including 1,931 authorities to the British Library. Previously we had a bespoke script which would collect locally-saved records from various staff PCs, combine them into a file, and send the file via FTP (a frequently error-strewn process requiring significant ICT input). With Alma we add a simple textual flag in a local field, allowing the records to be collected by a saved search set, which in turn is published via a saved publishing profile. A process which used to take half a day can now be completed in minutes, a significant time saving allowing staff to concentrate on meeting record creation targets.

Another significant change since our implementation of Alma is that we have decided not to create and maintain local authority records. This has had a particular impact on our non-print cataloguers, who were used to undertaking quite detailed internal authority work, but this effort can now be channelled into increasing our contributions to NACO, ensuring greater benefit. The previous authority data has all been retained for future re-use. We have some work to do still in analysing the results of the automated matching of headings and authorities, but staff are very excited about the possibilities offered.

As a systems librarian, my personal favourite aspect of Alma is the power it provides to manipulate and correct data. Having over the years worked on migration of several legacy datasets, and on trying to mitigate the discovery problems often inherent in legacy data, I was used to needing to write detailed mapping documents to be scripted by developers, and endless testing of the data outputs. On Alma, utilising sets, indication and normalisation rules and running jobs, I can personally implement data changes and corrections that would have taken weeks previously. Already we have been able to resolve long-standing issues using these tools, and I look forward to being able to systematically improve both the quality and discoverability of all our records in the future.
We went live with Alma in July 2015 and although it has been an up and down experience it has proved to be a useful learning curve for us in the metadata team. Two areas, in particular, that have changed for us are bibliographic record Import Profiles and e-resources.

In Voyager our team worked in a bit of a bubble. We retrieved individual records, did what we had to and then saved them, and if they merged strangely or not at all we didn’t worry much. Creating and using import profiles ourselves means we are more aware of what happens either side of our part of the acquisition process now. We can and do identify and often solve problems ourselves in Alma because we know what an import process should be doing to a set of records, and because of the reports received for each import.

This means that we are able to be more proactive. I have been able to add processes to import profiles (although Ex Libris say we are unable to add multiple processes as yet) which saves time in running jobs to sets of records after import, for example. 0.5 of my role is in Open Access and I have also been able to use import profiles to import open access book collections and acknowledge vendors such as Open Book Publishers – something I couldn’t have done before. Vendor names also allow us to much more quickly resolve user e-book issues because quite often we have several differently supplied versions and need to differentiate them in the user interface.

In the past we haven’t been very involved in e-book imports and description – records do come via us now and we have a say in whether or not we use Alma’s Community Zone (CZ) records. With a collection like Proquest Elibrary (formerly Ebrary) for example, we couldn’t keep on top of new and withdrawn books when we hosted records locally. Being able to link to dynamic Community Zone (CZ) collections was part of Alma’s attraction for us and the move has been effective. However (!) the CZ records are often very poor so we see our role now as advocating collectively for better quality metadata across the CZ so we take part in WHELF discussions between vendors and Ex Libris and we see more of this work on the horizon.
Finally, working closely with e-resources colleagues to create a new e-book collection has taught us a lot about how to collaborate to get the best out of Alma. Coleg Cymraeg Cenedlaethol (CCC) - E-lyfrau is a Welsh language collection that has been digitised by CCC and is available via their webpages. We wanted to have the e-books available through Primo so have created records and portfolios and have started the process of having the whole collection available to all Alma users through the CZ. While doing this has been fraught with strange problems – we have already made the collection accessible but then couldn’t add any new portfolios/books, for example—it has been a useful learning experience that we could never have tackled before; perhaps a good summary of our Alma experience so far!

Primo screenshot of the E-lyfrau : CCC collection level and individual e-books
I started using Alma in March 2016. From the outset the features that grabbed my attention, opening up long vistas of cataloguing possibilities, were the advanced repository search & creating and managing sets of records. I used the advanced repository search extensively during data migration testing as it provides a method of querying titles, (physical, electronic and digital), holdings, items and authorities (if you are managing these locally), in real time.

It’s not as granular as Analytics reporting, which I have still to get to grips with, but it has been invaluable for identifying post-migration data cleansing tasks e.g. items with no barcodes or no item policy; holdings with no items; bib records for print resources containing 856 tags; bib records with resource type “undefined” etc.

Getting to know our data has helped me to establish cataloguing housekeeping and collection management processes and routines.

We are weeding rigorously ready for RFID implementation so monitoring withdrawn items and suppressed bibs and items required for reclassification is important. There are many ways to query the data and you can then save sets of records for further analysis or bulk processing. Once created the sets can remain private or be shared. They can be logical (real time i.e. search results are refreshed each time that you view the results) or itemised (static list of records.) You can edit sets adding more conditions or combine sets but I think that my favourite feature is the ability to choose to catalogue a set of records. This option moves the set contents into a separate folder in the Metadata Editor ready for further work. So suddenly, combining the advanced repository search and mange sets functionality, it was possible to identify and re-catalogue artist's books that had been purchased prior to the creation of our Special Collections.

The possibilities for reviewing and editing records are endless!
These are some of the practical benefits that Siân, Amy and I have realised within our respective institutions but implementing as part of a consortium presents many exciting collaborative opportunities.

**WH ELF Consortial Cataloguing Opportunities**

We are now part of a group of cataloguers using Alma so we can forge closer relationships, share expertise, devise standards and hopefully reinvigorate cataloguing training activities in Wales. The first meeting of the WHELF Cataloguers Group took place in November 2015, even before all the consortium partners were live with the system. We were briefed by ExLibris about the creation of a WHELF Network Zone which would allow us to share and contribute catalogue records to a centralised database. There is support in principle for this but not at the risk of jeopardising our metadata. There may be other initiatives that we can tap into to realise our ambitions!

Collectively we anticipate quicker cataloguing of standard materials hopefully releasing time for cataloguing special collections & donations; and metadata enrichment projects e.g. for legacy data, institutional repositories. Reclassification project work is highly desirable to ensure that the virtual browse functionality in Primo can expose more of our collections. It might also be possible to found a NACO funnel for Wales to improve our authority control and open up some exciting opportunities for CPD.

We also collected data on cataloguing tools and services in use across the Consortium; the number of cataloguing staff in each partner institution; special cataloguing expertise e.g. languages, formats. The meeting resulted in the creation of a Yammer community where we have continued to ask questions, comment on documents and share ideas e.g. examining the quality of records for eResources in the ExLibris Community Zone and working with the Company and our vendors to improve the search and discovery experience in Primo for our users.

It’s possible that we could share the costs of purchasing library resources and services. For Cataloguers this might mean negotiating discounts for cataloguing tools e.g. RDA Toolkit subscription or records for copy-cataloguing. Perhaps enrichments for Primo or even Training/CPD to ensure that there is a skilled workforce to create, assess and manage metadata now and in the future.

We hope to agree cataloguing standards for AACR2 and RDA records, as well as how we handle hybrid records. We need agreement on the use of local tags across the Consortium (important for a cross-WHELF search in Primo.) Then there’s the creation and maintenance of Library of Congress name and subject access points & the use of other specialist vocabularies e.g. AAT (Art & Architecture Thesaurus.)

One of the more exciting opportunities is the possibility of developing generic cataloguing training for use across WHELF based on best practice and agreed standards. We would also like to incorporate accredited cataloguing training available via distance learning from UK Library Schools or possibly organisations based overseas e.g. ALCTS, Library Juice.

So exciting times ahead!

Roles & responsibilities may change but we believe that consortial cataloguing and the anticipated freedom to embark on new and exciting projects provides many new opportunities for those willing to grasp them.
Introduction

My poster presentation at the CILIP CIG conference was an overview of a research proposal which I created as part of my postgraduate Information Management course. It examined how metadata can be used in cultural heritage organisations to denote the historical and cultural context of heritage objects and documents. It brought in strands of thought from archive and museum studies alongside issues surrounding specific technological implementations for metadata management. The literature revealed the lack of standardisation in this particular area of data curation and the difficulties of creating ‘one size fits all’ approaches in an area where every collection and archive has its own idiosyncrasies and where the information needs of users can be very different. This is a vast field of study, and the proposal gave a very broad overview of ideas and methods which have been used in the past and suggested potential research questions for practitioners about what they valued most in contextual metadata.

Since then, I have been involved as a volunteer with an archive project helping to arrange and catalogue a small collection at Guildford Cathedral, with the ultimate aim of providing access for the public via an online catalogue. Given my background reading on this topic, it was the perfect opportunity to get involved in the creation and use of heritage metadata in a real-world setting. In contrast to the broad, theoretical view of metadata management, I discovered that the emphasis within projects such as this comes back to the question of accessibility for the user. The often limited time, resources and funding available mean that metadata solutions must be efficient but above all fast and relatively cheap to implement. Furthermore, it often falls to volunteers to sustain the collection once the initial setup project has been completed, therefore solutions must also be user-friendly and easily maintained by non-expert staff. This article aims to re-examine some of the themes of my original research proposal in terms of these realities, in order to give a new perspective on metadata creation and curation within cultural heritage collections as experienced by users, archivists and cataloguers ‘on the ground’.

Digital Surrogates and Web Access

Integration of collections with some kind of online presence is very often one of the key aims of heritage cataloguing projects. Indeed, in any modern archive, an important part of collection management is the creation of digital copies of documents which can be used for reference and preservation, or for inclusion in promotional materials and Online Public Access Catalogues (OPACs). This requires metadata to be created for three tiers of documents:

- the original objects, be they photographs, books or artefacts
- a high-quality digital surrogate for reference, preservation (in case of accidental loss or damage) or to enable fragile or degraded items to be examined without need for further handling
- one or more lower-quality digital images intended for use on websites or in literature to provide information or promote the work of the archive.

These discrete documents all require different metadata which reflects their separate uses but which is also interlinked to show their relationships with each other. For example, a document’s OPAC entry may contain a low-quality digital surrogate whose metadata may point to both its location in the file plan and the location of its high-quality partner. The high-resolution image is normally kept for internal use, or in some cases for use by researchers specifically requesting access.
Physical locations of artefacts can also be related to the virtual location of their digital surrogates via internal metadata. This arrangement can be seen as a kind of merging of collections metadata (e.g. the OPAC) and records management metadata (e.g. the internal digital file location). The two structures may be similar or different depending on the particular nature of the collection or the storage needs of the organisation. The important point is that the line between organisational records and collection metadata is becoming more blurred in modern collections, which may have potential implications for metadata management in the future.

Automatic Metadata Translation and Extension

Within online OPACs, technological enhancements can be used to automatically augment metadata to aid both cataloguers and users. One example which has been used in the Guildford Cathedral archive concerns the recording of dates. Researchers coming to an archive or collection catalogue may have an exact date for a particular document, or they may have nothing more accurate than a decade or even a century. A catalogue or resource can be made more accessible for these users if the implementation of the metadata takes this into account, particularly as an archive will tend to have more accurate information on a document than an average user. This can be achieved via the use of automatic metadata translation, which can be implemented very easily in date fields. Using this method, a document catalogued under the date “1957”, for example, can also be automatically listed as “1950s” (and by extension “1950-1959”), or even “20th Century” or “1900-1999” if appropriate, with no extra cataloguing work needed. The process also works in reverse, so that a user searching for an exact date will be able to find a document catalogued as “1950-1959” by searching for “1957”. Cataloguing under terms such as “circa 1957” will automatically add metadata to the document for dates three years either side. This programmable metadata translation could even be tailored to the needs of a particular archive. For example, if there is a significant watershed date around which documents fall, the system can take this into account, for example making “pre-1957” and “post-1957” a part of the metadata translation. This feature can be implemented in many OPACs and can save time and resources for cataloguers, as well as increasing potential routes of access to documents.

User-generated Metadata

In some ways, selection and implementation of metadata is no longer the sole domain of archivists and cataloguers. Many collections will allow users some limited access to metadata creation in the form of keyword tagging. However, users are less likely to follow controlled vocabularies, which can lead to problems in terms of the accuracy, completeness and consistency of the metadata they create. Take, for example, the many ways in which “World War Two”, “Second World War”, “WWII” etc. can be conflated. For these reasons, user-generated metadata is often implemented as a ‘bolt-on’ aspect of the final web OPAC and is unlikely to become an intrinsic part of the back-end archival description. Yet, it can be a crucial part of the public engagement aspects of a collection, allowing users to make connections or observations about the documents in their own words. In this way, allowing this kind of sandbox area where ‘bad metadata’ can exist may be a legitimate way of providing avenues of access which may be otherwise closed to some users. In some instances, it may even act as a two-way process of communication if, for example, a user has personal knowledge relating to an archive document. Carefully managed, this can be a crucial way of drawing in users and augmenting a collection with knowledge which would otherwise be lost.

Hierarchies for Context

The most important function of metadata and archival description – context – must not be forgotten. In small collections and in time- and resource-limited projects, it is often more effective to denote context purely via a hierarchical structure. This is used efficiently in the Guildford Cathedral archive to delineate the separate strands of documentary provenance. For example, the photograph collection has a separate hierarchy to the diocesan records though there may be instances where the two collections overlap. Metadata standards such as ISAD(G) can then be used to provide further granularity and cross-referencing by ensuring that specific information is included in each record.
The popularity of the ISAD(G) standard reflects its ease of use and the relevance of its elements, yet in some ways it is the hierarchical structure which provides the most helpful context. It can give an immediate overview of a collection, showing the different aspects at one glance. It also provides a ready-made method of describing the physical location of documents, as items can be kept together depending on where they sit in the hierarchy. This aspect can too often be neglected in the rush to digitise archives, and it is also useful in any ongoing project where documents are constantly being taken in and out of the archive. In this way, the hierarchy structure itself becomes a valuable metadata tool which carries a lot of information very efficiently, even before drilling down to the item-level metadata.

Conclusion

This overview of some of the metadata methods and concerns within small heritage projects brings home the message that the user is now very often the focal point to the organisational decisions made for the preservation of collections. Ease of access, new ways of encouraging engagement and the use of digital images make for a constantly changing environment in archives and collections. Underlying this, however, are still established methods and approaches adopted by information professionals about how metadata should be used, as well as technological considerations about how to ensure access and save time and resources. In terms of real-world projects, it is clear that this will always mean striking a balance between creating the best possible information structure for a given collection, and working efficiently in short timescales. Whilst it is important to examine metadata standards and how they aid description, too much preoccupation with finding the 'perfect' standard is perhaps unhelpful. In this sense, my experiences within a real archive project corroborate many of my findings in the literature – namely that heritage organisations are constantly innovating and looking for a combination of solutions which allows them to best serve the needs of their collections. If quality and provenance of the archive information can be best preserved through using more than one metadata management method, then these needs are being served.

I would like to thank the People’s Cathedral Project team at Guildford Cathedral for giving me the opportunity to experience a cataloguing project first hand. More information can be found at http://www.guildford-cathedral.org/about/the-peoples-cathedral.

Further Reading


Overview of Kingston and ordering

Kingston University is a former polytechnic with around 20,000 students and 2,000 staff across 5 sites. The Learning Resource Centres (LRCs) at 4 sites are staffed by a FTE of 96 staff. We have 17 staff in Collections and our Library Management System (LMS) is Alma (we went live in June 2014).

Our subject teams place orders and are responsible for selecting records and if possible, to specify a classmark. Most items arrive shelf-ready. On arrival the items are receipted by the Acquisitions strand of Collections. The items are then shelved or passed to Cataloguing. Items needing cataloguing or classification are identified by a note placed by subject teams when ordered.

Gathering data

Prior to 2013 we didn’t track throughput. We do have a KPI (Key Performance Indicator) around the percentage of books that must be on the shelf within 1 working day of receipt, but other than that there’s no target as to how many books we must catalogue in a day/month.

In mid-2013 we considered purchasing direct from the vendor platforms for our 2 main suppliers, rather than placing orders in the LMS. As part of the preparation we started to look at where we were getting information from when cataloguing items (whether we could download a new record or use e.g.: WorldCat for information). We noticed a lot of items were coming to Cataloguing unnecessarily and wanted to get a better idea of why this may be.

From February 2014 we recorded:

1. Source of record / information (for cataloguing)
2. Records only needing a classmark (no cataloguing required)
3. Only minor fixes (ie: typos)
4. Didn’t need to come to us
5. Duplicate records
6. Reclassification (either requested or from a Dewey 23 reclassification)
7. Item fixes (i.e: changing the loan period)

---

1. We did do this, but abandoned it 6 months later due to a variety of reasons.
2. Not all data is included here. For further information please contact me.
3. Items with multiple things ‘wrong’ i.e: 2 different records at 2 different classmarks…
Figure 1, above, shows how records were completed when Cataloguing dealt with them (not at point of order). Three things to note:

1 – Added British National Bibliography (BNB) as a record source in Aleph, the impact of which cannot be underestimated – in one afternoon my colleague cleared 3 shelves of books!

2 – Move to Alma with a wider range of sources for records - noted as a measurable benefit of moving to Alma.

3 – In October we started to get items coming in that had been ordered in Alma rather than in Aleph. As these often had a record, we could focus on more specialist cataloguing.

Not all items handled were catalogued (Figure 2, below). On its own these figures don’t reveal much. Figure 3 (below) shows in detail what these other interventions were.
By the end of the year more items only needed a classmark or didn’t need to come to us. Whilst this justified any grumbles about “Why on earth did this get passed to Cataloguing?” we needed to look at it in more depth.

Into 2015:

In 2015 we focussed on 3 key areas.
- Items only requiring classmarks – grouped by faculty
- Duplicate records – to try and trace when the extra record appeared.
- Didn’t need to come to us – noting down reasons.
Each faculty had, at some point, an item that just needed a classmark (Figure 4, previous page), the majority though were in Art and Design and Education. This fitted with what we thought and was largely because we believed vendors that supplied those books didn’t do shelf-ready.

![Fig 5: Items with duplicate records - 2015 - breakdown by Faculty](image)

The numbers of duplicate records (Figure 5, above) was much lower than anticipated. With Alma we can see who placed orders and could approach them to see where the confusion lay.

The most interesting results, to me, were always going to be the reasons behind items being passed to us that didn’t need to be (Figure 6, below).

![Fig 6: Items that didn’t need to come to us – 2015 - breakdown by type](image)
The main reason was items marked as ‘Pass to Cataloguing’ at point of order, despite having a good record and classmark specified, indicating an area for further training.

So what now?

Classmarks and workflows

We realised we could specify classmarks for everything ordered, regardless of whether vendors could supply shelf-ready or not. Most of our items are fairly simple to classify, the ‘Note to vendor’ field is visible to acquisitions staff and having the note for non-shelf-ready vendors doesn’t cause any problems. We retrained all staff involved in ordering and provided guidance in selecting classmarks.

Discussions with another library revealed the main supplier we used for Art and Design items could provide books with spine labels and, after discussions with the supplier, this was instigated.

We took the opportunity to evaluate how we handled items arriving in bulk from a supplier where the orders are created on arrival. These were for Teaching Resources Collection – a mixture of books and physical resources such as puppets or musical instruments.

The old system:

Items arriving from suppliers were passed to the subject team to download records and add orders. The items were then passed to Acquisitions for receipting and to us for cataloguing/classification. They were then passed back to Acquisitions for labelling and to the subject team for information.

This involved a lot of double handling and as items often are in sets, it’s much easier to catalogue an individual item and duplicate the record than to edit multiple records. In March 2016, a revised workflow was put in place.\(^4\)

---

\(^4\) This workflow only applies to purchases requiring orders on arrival. Everything else still goes to Acquisitions on arrival.
Items go straight to Cataloguing and staff download/create catalogue records and apply classmarks. The item is passed to Acquisitions who create the order and receipt it, prep the item and pass it to the Subject team and then the item is shelved. Overall, it's been a positive improvement.

**Staff training**

A simple document showing acceptable/not acceptable AACR2 and RDA records was produced and individuals were approached if acceptable records were being marked as ‘Pass to Cataloguing’. I’ve been keen to stress I’m not constantly checking who does what, just following up where there are knowledge gaps to be filled.

For example, a reason for passing items to us is the presence of a 263 field (indicating pre-publication records). Speaking to a colleague who had passed a lot of items to us incorrectly, it transpired they thought this applied to items that had any field beginning 26-, so all the RDA records with 264 (publication information) they thought were ‘unacceptable’. I can see where they were coming from, but it had never occurred to me that this would be a cause of confusion.

I’m trying to foster a culture where it’s ok to ask questions as it’s much easier (and cheaper!) to answer a quick question in advance than to unpick something when it’s discovered, maybe months later.

**Where are we now?**

So, has all the work paid off? In terms of items only requiring Dewey numbers, it certainly has (Figure 7, next page).
The number of books requiring only a classmark has dropped to almost none, but as new procedures were only implemented in April, it’s likely there are still some older items on order. There will also always be one or two items where the item has to be viewed to decide where best to put it.

The graph I was most looking forward to seeing was the one for ‘Items that didn’t need to come to us’. I was hoping that it would be on a similar level to Figure 7. However, my heart sank when I saw the numbers increasing! What was going on?
Because we had been recording the reasons for these items coming to us, it was possible to look into it in a little more depth (Figure 9 below).

![Fig 9: Items that didn't need to come to us – reasons Apr-Jul 2016](image)

Whilst a few were still being incorrectly marked as 'Pass to cataloguing', there were other things going on. Items 'Not marked as anything' has always been an issue. Unfortunately, as good as cataloguers are, we can't see inside a person's head to know what they were thinking when they passed the book to us. These few may have simply been human error. The largest area though were books that had been classified but had no label on them. These were items with thin spines, where we normally label them side-on. The supplier can't do that so had simply thrown the label away. We've now briefed Acquisitions staff on how to handle them and find the classmark.

I also compared April – July 2016 with the same period in 2015 (Figure 10, below). This showed the balance was moving towards problems with the supplier and away from items being marked incorrectly. Furthermore, the number of items with 'other' issues had dropped, suggesting people are more aware of how to handle queries on items.

![Fig 10: Items that didn't need to come to us – reasons Apr-Jul 2015 v 2016](image)
To have a slightly better looking graph I redid Figure 8 to remove the items with the ‘stickering’ issue (as that will hopefully not reoccur) (Figure 11, below). It does look a little better!

![Graph showing varying numbers of items coming in during the year and comparing month on month doesn’t always give an accurate representation. Therefore, I’ve taken a monthly average for each year for the items that didn’t need to come to us.](image)

Fig 11: Items that didn’t need to come to us (2014-2016) minus sticker issue

The graphs show the varying numbers of items coming in during the year and comparing month on month doesn’t always give an accurate representation. Therefore, I’ve taken a monthly average for each year for the items that didn’t need to come to us.

- **2014 – 27 items per month**
  - 29% drop

- **2015 – 19 items per month**
  - 21% drop

- **2016 – 15 items per month**
  - 62% drop

(The 10 items as opposed to the 15 for 2016 was removing the stickering issue).
Whilst 17 items a month might not sound a lot, on the busiest month in 2014 it took around 2 days to deal with all these unnecessary items. As we have multiple sites, items can sit for a few days before they are looked at. Our roles are, as with many people, always changing. I now spend a great deal of my week supporting the Research Repository librarian with Open Access, so time is more precious. Additionally, in the new Town House (new LRC), the Collections staff won’t be permanently based there (other than the Archivist). Reducing time spent there unnecessarily will therefore be important.

What I’ve learnt

Most importantly, I think this project has given us in Cataloguing a chance to have a dialogue with the subject teams and for both sides to understand better what they do. We’ve been able to explain why the throughput rate is important to them – i.e.: effect on usage of stock or satisfaction of students.

It’s important to take the time to question practices and to investigate the reasons behind any trends you observe. Get others involved remembering to couch issues in terms of what’s in it for them but be aware of what they actually know, not what you think they do! Above all, don’t be afraid to change/stop things if they’re not working and be persistent - change can take time and several attempts before you get what you want done.

I’m more than happy to answer any queries on any aspects of what we did – my email is k.clifford@kingston.ac.uk
It sounds so simple … “Let’s go shelf-ready with the AV”. After all, we’ve been acquiring print materials shelf-ready since 2003, it works perfectly (fingers crossed), what could possibly go wrong?

At Leeds Beckett (then Leeds Met) we were early adopters of shelf-ready acquisition in the academic sector, and on the whole it runs very smoothly. As we went more fully shelf-ready with book stock (and some older members of staff retired…) I was able to chip away at the last bastions of local practice until we reached the position that for most print material we now produce and accept standard AACR2/RDA MARC21 records, with (very nearly) standard LCSH and DDC22 or 23. Cataloguers however have not become totally redundant, though in that period the numbers have been reduced from 2.5 fte (with extra help when needed from management and other librarians) to 1.4 fte. That’s me, and part of my colleague Tom, when he’s not trying to manage the ebook packages (also known as trying to herd cats). The work is more focused now on quality control, upgrading our older data, and dealing with things we can’t manage shelf-ready (theses, some foreign language publications, out of print and, originally, all the AV - not to mention the e-book packages with their MARC record sets of varying quality! Large language teaching sets had to be taken out of the programme, because of the tendency for sets to be split by DDC. Similarly our School Practice material was also taken out of the project because we had to reclassify the Primary material from the 370s to the wider schedule.

We’d been asking our suppliers for records for AV material (mainly films on DVD, music on CD, and computer games) for some years, and they had been unable to provide quality records, or customised processing. In particular, no one seemed to be supplying the LCSH and DDC that we need. The AV shelf-ready plan was kick started by one of our major suppliers who suddenly offered us good standard RDA records, and were willing to apply the appropriate class and subject headings, without which it wasn’t worth us buying records in. The project was led by the Acquisitions team but required a lot of input from other areas – subject librarians, collections maintenance, and of course cataloguing/metadata (me). It was also the first use in the Library of a new University-wide formal project procedure, with such success that the project is now being used by the University as a case study of how the procedure works.

We decided to start with DVD feature films – a large part of our intake, and fairly straightforward. My first task was to produce a specification for the records – we try to keep that quite simple (MARC21, AACR2 or RDA, DDC 22-23, LCSH, MARC8 ) and aim at sorting out details when we get sample records. In fact the records were fine, with only a few minor queries. The physical processing was a bit more of a problem – 9 different types of material and 12 different processing actions. There turned out to be a great many possible permutations! At that point the supplier’s promises of customised processing started to look a little shaky…. They went backwards and forwards and finally produced a spec which worked for both sides – this may have been the most difficult part for the acquisitions team.

I also had to come to a decision about the filing letters. Our historical policy for AV had been to file items by title – everything non-book was catalogued in-house with title main entry. The policy predated my arrival, and, I think, was intended to save time and effort on deciding main entry, especially with non-expert cataloguers not used to processing music and films. With someone else now being paid to make the decision, we decided to use standard main entry, and thus filing letters could now be from 100, 110 and 111 fields, as with most of the rest of our materials.

This of course meant that we’d have double sequences on the shelves, which we would have no time to rectify, but I made a fairly arbitrary decision in favour of the change. One argument for it was that we would have a double sequence for films anyway. Until 10 years ago the policy of the cataloguing department was to keep things very short and simple, especially for AV materials.
It was impossible to tell different versions of the same film apart from the bibliographic record, non-LCSH subject headings had been applied seemingly at random, and a local variation of DDC had been used to comply with the former Arts Librarian’s preferences. (she had preferred films to be classified and shelved by “country” – she was neither clear nor helpful on whether she meant country of production, filming or language…). I had to produce some notes explaining the changes and their projected outcome to the subject liaison librarians, but I think I may have blinded them with technicalities and jargon, and they agreed without argument. The argument clincher as always was the promised reduction in throughput time.

In the meantime, the Acquisitions staff were battling with problems caused by Coutts’ need to use EANs in the order process, when we often had ISBNs, or other standard numbers, or Amazon numbers supplied by the librarian ordering the item. The limitations of our LMS, and the wish to streamline the order process and keep it consistent with ordering non AV stock, meant that in the order forms, the EAN had to go into the box normally used for an ISBN, with the result that it then appeared in both the control number and ISBN field in the record. We decided to retain the EAN as a control number, while removing the “i” prefix that denotes an ISBN, and to make putting it into the proper field in the bibliographic record part of the loading process done by Acquisitions staff – a small fiddle, but it works. All of this of course required more staff training, and more re-writing of manuals and guidelines. Boxed sets were taken out of the programme, because of the difficulties of shelving and physically processing them, as were DVD/BluRay combinations.

Eventually we received the first test records and DVDs. We’d decided to start with the feature films, and then to expand further – probably a very wise move. A specific Acquisitions Assistant with an interest in film and music had already been responsible for much of the ordering and processing, and I trained him to check incoming records, and edit them where necessary. He’s since rolled out the training, but remains the local expert. There have been odd problems with punctuation in 245 fields – I think it’s being lost in transit in the ether somewhere! – but nothing too terrible.

Unauthorised headings are picked up by a weekly report, as with the non-AV items. These are then checked and authorised, by the cataloguers. The only DVDs that now come to a cataloguer are those with unauthorised titles (including series), which I prefer to see on the item itself, and queries and problems. And of course, those that we can’t order because we can’t provide an EAN – a diminishing number, mainly of strange arty materials!

The double sequences on the shelves remain, but the material doesn’t have a long shelf life – it is either worn out, damaged or “walks”, so the problem is self-managing. The change from the old non-standard subject headings to proper LCSH seems to have been regretted, or even noticed, by no one. I confess to retaining the old local headings containing the word “Films” and not changing them to “Motion pictures”, on the principal that a) “Motion pictures” is horrible, and b) there are far too many of them, and c) the Library of Congress will surely get round to changing it sometime themselves. Please????

We’ve recently moved on to including music CDs in the programme, and again, it’s going smoothly. We do still have changes to make in our practice and policies (for example, our record format for all AV material is VM, whether it is audio or visual. I have no idea why: and we need to change this as it means we can’t make proper use of the fixed fields, which are becoming more important as some of that information is used by the Discover system. This is a problem for the Systems Librarian though, and fairly low on her list of priorities!).

The end result was an enormous improvement in customer service: a system which integrated smoothly with our print and e-resource purchasing: which didn’t negatively affect processes in other areas of the Library: and which speeded up the throughput time for audio-visual resources from up to 8 weeks to under 3 days. This is despite the reduction in cataloguing manpower by more than half since our first shelf ready experiments, 13 years ago.
The project re-emphasised for me the importance of being prepared to take pragmatic decisions about cataloguing matters when it is necessary. As a dyed-in-the-wool cataloguer, I would like our shelves, our classification, filing and subject indexing to be perfect: however this is just not possible, especially with our reduced resources. We need also to compromise where necessary with suppliers, and to abandon old procedures happily when the rationale behind them is no longer valid. Finally, none of this could have been achieved without the help of the Collections and Acquisitions staff, who have readily taken on new responsibilities and acquired skills.
Abstract

In October 2015 the British Library published its Collection Metadata Strategy. This paper reviews the reasons for developing a metadata strategy. The paper will focus on the challenges (familiar to most libraries) that have prompted the British Library to articulate a holistic vision of its collection metadata as a key corporate asset, including:

- The extension of legal deposit to non-print media and the growth of bulk digitisation have created demands for more metadata and more flexible metadata. Metadata is no longer only for discovery, it is also critically important for the automation of workflows and processes.
- The switch of content from physical to digital carriers challenges distinctions which are embedded in our workflows and the metadata standards we use.
- Changing government policy and user expectations that challenge traditional business models for dissemination and re-use of metadata.

The paper will include an update on progress achieved so far and consider the lessons learned to date.

Introduction

The British Library is the National Library of the United Kingdom. It is one of 5 Legal Deposit libraries. There is a legal requirement for publishers to deposit publications with the library within one month of their publication; since 2013 the library has been empowered to receive non-print publications, including digital resources and to archive the UK Web domain. The Library was created by Act of Parliament in 1973 by merging several existing libraries and agencies, including the British Museum Library, the British National Bibliography, the National Lending Library, the Patent Office Library, the India Office Library and Records, the National Sound Archive and most recently the Public Lending Right. Each of these institutions had its own collections, housed in its own building, curated by its own staff and with its own catalogues and databases. With the passage of time, most of these historic silos have been eroded and the Library’s assets, including the collection, buildings and staff, brought under centralised management.

Why a collection metadata strategy?

Convergence

The implementation of Aleph, the Library’s first true Integrated Library Management System, in 2004, resulted in the migration of approximately 13 million bibliographic records from multiple catalogues and bibliographies into a single production database. However this undertaking only integrated metadata for published resources, including books, sheet music, maps and serials. Metadata describing manuscripts and archives, sound recordings and journal articles, which now total well in excess of 40 million records, were not in scope for Aleph and continue to be managed in separate repositories.

- Aleph 500 ILS for published resources
- ETOC for articles (In-house database)
- IAMS for Manuscripts and archives (in-house Integrated Archives and Manuscripts System)
- SAMI for Sound and Moving Image (SirsiDynix Symphony ILS)

Each of these repositories constitutes a distinct silo, with its own standards, workflows and associated staff. From an operational perspective this is inefficient and inflexible; for example, staff cannot be easily transferred between workflows to relieve short term bottlenecks.
More seriously, this leads to fragmented services, including a complex discovery layer, which in turn makes it difficult for users and staff to understand the collection and find the resources or information that they need. Many of the platforms we use to create, store and disseminate our metadata are reaching end of contract or end of life and may need to be replaced. So there is an opportunity to take a more holistic view of the functions our metadata needs and develop a more coherent architecture.

Hidden metadata=hidden collections

The user may also assume that everything we have is recorded in our catalogues. This is not the case. There remain substantial pockets of “hidden metadata”. Although most of our catalogues of publications have been converted to MARC there are still about 2 million catalogue records to be converted, split more or less evenly between publications and archival/Mss collections. In addition, there are substantial backlogs of uncatalogued materials.

Changing expectations

Our legacy metadata is being asked to do things that it was not designed to do. For example, when Panizzi began work to create the British Museum catalogue, he did not think it necessary (in most cases) to record the language of the content, because this would be self-evident from the description. When these records were transcribed and retrospectively converted to MARC there was therefore no information to put in the MARC 21 008/35-37 language bytes. Consequently, if someone wants to know how many books the Library has in Arabic or wishes to create a visualisation of the collection by language or create a data set of records for resources in European languages, comprehensive answers cannot be provided.
The preceding examples are also indicative of the growing significance of collection metadata as data: raw material for new research or management information. The Library has traditionally provided bibliographic records to libraries, publishers and other third parties in MARC but the technical difficulties of using MARC data put them beyond the reach of all but the most committed researchers. There is also increased interest in sharing metadata for other types of resources, including archival collections and manuscripts. Central government has become increasingly interested in the economic potential of digital information, including metadata, and public bodies are encouraged to open up their data for reuse.

The digitization of content erodes many of the practical justifications for historic distinctions between collections and users reasonably expect to be able to search holistically, regardless of content or carrier types.

In 2013 legal deposit was extended to non-print media. Non-print media include digital books and journals, but also the UK web domain. Processing digital resources creates new challenges of scale, granularity and volatility. Our expectations were that as publishers were switched from print to electronic legal deposit the increase in the intake of electronic resources would be balanced by a reduction of the print. In fact, there has been little fall off in print but even with only a small number of publishers switching to digital, the digital intake for 8 months was equivalent to about half of the print intake for a whole year.

![E-Books Received 2015/16](image)

**Figure 2: Legal Deposit E-books received May 2015 Jan. 2016**

It is clear that once the digital tap is open, pretty much everything the publishers issue will flow through it, including back catalogue:
and electronic publications with no UK imprint:

![Ebooks by Year of Original Publication](image)

Figure 3: e-books by year of original publication received by Legal Deposit 2015

From a national collection development perspective this is a good thing; from a collection processing perspective a 50% increase in intake within 8 months poses an interesting challenge.

**What is the collection metadata strategy?**

Collection metadata is defined as structured data that supports collection management, preservation and accessibility. Metadata recorded by the Library for other purposes is out of scope. Collection metadata should be managed as an asset, like the collections or the library staff and estate.
The strategy provides us with a framework to manage that asset better and to unlock more value from it for our users, our staff and our stakeholders.

“Our vision is that by 2020 the Library’s collection metadata assets will be comprehensive, coherent, authoritative and sustainable, enabling their full value to be unlocked for improved content management, greater collaboration and wider use of the collection.”

If we unpack this further:
- **comprehensive** means that we should have metadata for everything in the collection;
- **coherent** means that silos will be broken down to provide the ability to search across the whole collection;
- **authoritative** means that the metadata we create is accurate and trustworthy;
- **sustainable** means that we are looking for long term solutions to long standing problems quick wins.

To realise these aspirations we set the following broad objectives

- Drive efficiencies in the creation, management and exploitation of collection metadata to support delivery of the Library’s strategic priorities and programmes
- Improve the Library’s return on investment in its collection metadata assets by ensuring their long term value is maintained for future activities
- Open up more of the Library’s collection metadata to improve access to Library content and promote wider re-use

To manage delivery of the objectives we set out a five year plan, which is flexible enough to respond to changing needs and priorities. The plan identifies the strategic priorities under each of these broad objectives and sets milestones and targets for delivery.

---

**Figure 5: Spider diagram illustrating strategic plan, 2015-2020**
In addition to the deliverables identified by the plan, there are performance indicators to measure progress. The strategy is aligned with the British Library’s corporate strategy, *Living Knowledge* and collection metadata is fundamental to realising its aspirations.

**How did we create the strategy?**

Historically, responsibility for the Library’s collection metadata has been diffuse, with authority distributed across different collections products and services. A key element in the strategy has been to develop the Collection Metadata team, headed by Neil Wilson, as the acknowledged point of responsibility. This builds on the team’s well-established provision of MARC based bibliographic services by expanding its scope to responsibilities to include archives and sound recordings.

![Figure 6: Collection Metadata (high level) structure chart](image)

Each of these teams has a clearly defined area of responsibility, but they also work closely together to deliver solutions to problems. The Collection Metadata team is responsible for delivery of the strategy. A parallel structure has been put in place to provide strategic oversight and expert advice.
This means that collection metadata has a champion on the Library’s Strategic Leadership Team. Among SLT’s responsibilities is the review of new business cases, all of which now have to explicitly consider metadata requirements and impacts, thus ensuring that Collection Metadata is able to influence new initiatives. The structure also provides formal channels through which internal stakeholders are able to input to the development and delivery of the strategy and provides enhanced channels for communication.

The strategy was has been disseminated internally through a comprehensive communication plan. General information has been communicated through staff talks and our internal newsletter. The Collection Metadata wiki provides guidance about the services we offer and how to contact us. More focused information has been provided by engaging with specific stakeholders to develop policies and procedural documentation. The strategy is available to external stakeholders on our Website.

---

**Figure 7: Internal Stakeholders**

**Figure 8: External Stakeholders**
Where are we now?

Library's fiscal and planning year ends on 31st of March. The objectives for 2014/15 were completed and we are still on track, approaching the midpoint of 2015/16. The following progress summary is organized according to the priorities and is not intended to be comprehensive.

Strategic Priority 1: Drive efficiencies in the creation, management and exploitation of collection metadata to support delivery of the British Library's strategic priorities and programmes

- **New e-publisher metadata assessment & ingest processes**
  - Publisher and aggregator data is highly variable, even if it is received in a standard format such as ONIX. Collection Metadata has developed processes to assess the quality of metadata so that this can be taken into account when prioritising publishers for transition to e-legal deposit. Ideally, the publisher’s metadata will be compatible with the generic transformation. If not, a bespoke transformation may have to be developed using XSLT to convert the metadata provided from ONIX, Dublin Core, or csv, to MARC 21 or MODS.

- **Automated enhancement processes for e-books and Western European Languages**
  - Our long-standing “batch upgrade” process to match Cataloguing in Publication records with full records without cataloguer intervention has been adapted to create new automated workflows for WEL material (for which no CIP data is usually available) and e-books. This is still being tested, but we expect the processes to go live by November 2016.

- **Spread sheet data capture & crowd source workflows**
  - We have helped develop tools to lower barriers to productivity of staff on fixed term contracts. These include a spreadsheet based tool to guide input and support selection of authorised access points from NACO and FAST and XSLT transformations. This reduces the training overhead, because staff do not have to learn to use Aleph or other complex systems.

- **FAST consultation**
  - We have consulted on the possibility of using FAST more extensively in future as a potentially more efficient indexing system than LCSH.

Strategic Priority 2: Improve the Library’s return on investment in its collection metadata assets by ensuring their long term value is maintained for future activities

- **Documentation**
  - It is essential that we have a clear picture of our metadata assets. We have created registers of existing assets, including hidden metadata, to ensure that information about the metadata is accurate and current. This supports tracking of progress and other management information.

- **Metadata impact assessment for business programmes**
  - To manage workload metadata impact assessments have been built into the Library’s business planning process and we have recently instituted a work request system so that we can plan priorities and allocate resources more effectively.

- **Data enhancement**
  - Enhancement of the quality of metadata will be essential to realise the aspirations the library has to improve resource discover. Projects are underway to extend coverage of DDC and to improve identification of Works.
Strategic Priority 3: Open up more of the Library’s collection metadata to improve access to Library content and promote wider re-use

- **Free data**
  - In recent years the Library has sought to remove barriers to access to our data by exposing MARC data via Z39.50 and removing licence restrictions from non-MARC data.
  - The Z39.50 service is being used by more than 1550 users in 177 countries. It is particularly valuable to professionals in singleton posts and small institutions, such as school libraries.

- **Development of .csv “researcher format”**
  - We recognised that the formats in which metadata was available were themselves a barrier to researchers, who may not have the tools or expertise to interpret MARC or RDF. In 2015 we began to make metadata available in the so-called researcher format. The records can be imported from our downloads page in .csv (comma separated value) format. This format enables manipulation of the data using desktop applications, such as Excel or Open Refine. At present, the datasets are selected by collection metadata staff but we are testing a tool that would allow researcher to define their own datasets.
  - We have developed the capability to integrate data from both the printed collections and the MSS collection into the researcher format.

- **Linked Open Data**
  - We have worked with Fujitsu Ireland and TSO to improve our understanding of how the Linked Open Data BNB is being used. We want to know who is using our data. What data are they using? How can we optimise publication of the data?
  - We have enriched LOD BNB by the addition of 745,000 International Standard Name Identifiers. A project is currently underway to accommodate pre-publication information (CIP) within the model.

This implementation Roadmap for 2016 gives an overview of the activity.
Conclusion

Implementation of the strategy is still in the early stages. In addition to the progress already discussed, the strategy is already having a significant impact. The visibility of collection metadata has been significantly enhanced within the organization. This means that we are consulted much earlier in projects than would have been the case in the past and are able to influence rather than react to decisions. We have seen a substantial increase in demand for our input and we have introduced a work request system to manage it openly and efficiently. We have also been able to demonstrate the need for additional resource to deal with the increasing workload.

Establishing the need for a collection metadata strategy its articulation were difficult tasks that are now beginning to pay dividends. Specifically it has opened up the question of how we can take advantage of future systems replacements to converge our legacy metadata. Convergence of the metadata silos is a precondition for eroding the longstanding silos that hamper discovery of the collection and impede the efficiency of processing.

References

Unlocking the value: the British Library’s Collection Metadata Strategy, 2015-2018
Collection Metadata Strategy Roadmap, 2015-2018
British Library. Free Data Services
http://www.bl.uk/bibliographic/datafree.html
British Library. Downloads
http://www.bl.uk/bibliographic/download.html
What is it that cataloguers and librarians fear the most? I think it is unhappy users; users who cannot find e-resources within library catalogues and consequently cannot use these resources within publishers’ platforms.

As a publisher, Cambridge University Press does not want users to be unhappy because we know that unhappy users make: unhappy librarians. Librarians that cannot offer the best services to their users. And they can’t do that because of miscommunication from publishers. One of the many forms of miscommunication happens when librarians do not receive correct metadata from publishers, when this happens discoverability is seriously compromised. In order to have happy users and happy librarians, we know that we have to improve the discoverability of our resources and we know that supplying better metadata will enable us to achieve this. With this paper we want to show how Cambridge University Press is improving the discoverability of our resources, mainly through the improvement of our eBook MARC records.

This paper’s topics:

- MARC records and discoverability
- Improvements of the automated eBook MARC records at CUP
- Manual improvements of the eBook MARC records
- A special project: CLC eBook MARC records
- Special project 2: OPL MARC records
- Cambridge Core: eBook MARC records and KBART lists

MARC records and discoverability

A MARC record is the bibliographic description of a resource. And this description has to be properly coded and complete in order to achieve the best discoverability possible. What happens when this description is not complete? Here’s an example: I was looking for this eBook (Fig. 1) within the library catalogue of the University of Cambridge and I started my search as any student would have started: with a subject search. The results I got were: No results at all.
Why did this happen? This happened because the MARC record supplied for this eBook did not have subject headings. Consequently the record could not be found and therefore the eBook could not be found. As a publisher we do not want this to happen, because if I were a student I would think that the library did not have this title and then I would have gone to the librarians to ask for help. Consequently I would have wasted time and the librarians would have wasted time. This is a bad situation from our perspective, because we want our resources to be discoverable without any delays. We want to supply MARC records with data that enables us to achieve the best discoverability possible.

Improvement of the automated eBook MARC records at CUP

When I started working at the Press, I realized that our eBook MARC records needed to be improved, so I set out to change them in two ways. First we have modified the automated eBook MARC records, the ones customers can download from our platform Cambridge Core. In the following image (Fig. 2) you can see the improvements that have been made within these records.

On the left we have a record as it looked before. We changed this to remove AACR2 data, since we want to supply RDA records and we have also corrected technical errors and description errors. On the right we have a record as it looks now. It is a good record because is technically correct and properly described, so it can be uploaded to a library system without any problems, but it is still not a “perfect record”. It doesn’t have subject headings, the Dewey classification number and other important fields. At the Press we want to supply eBook records containing all the access points needed to achieve the best discoverability possible, so the second stage of changes was to manually improve our eBook records in order to accomplish this.
Manual improvement of the eBook MARC records

I have worked with different cataloguing tools: MarcEdit, to modify records, and also to merge our eBook records with the print LC records and also to create Macros in order to make these records (which had been created for books published in the 20th century so catalogued with AACR2) in RDA records; Bibliofile, which is a very good tool to use to examine and modify each record in detail; ClassificationWeb, a very useful platform supplied by the Library of Congress which I have used to check LC classification numbers, Dewey numbers and Subject headings. And the LC’s Authorities, which I have used to check the authorised forms of names and series’ titles.

On the left of this image (Fig. 3) there is the same automated record from the previous image, with no Dewey classification number, no subject headings and other elements. On the right there is a manually improved record as it looks now, after merging it with the LC print record.

I did this for the majority of our collections, Cambridge Books, Partner Publishers (previously University Publishing Online - UPO), Cambridge Histories, Cambridge Companions and Shakespeare Survey (about 24,000 title in total). After merging the eBook records with the print records, first I used macros I had created with MarcEdit in order to move data within the records so that everything was in its proper place, since these titles had been catalogued with different cataloguing rules. Then I checked that LC classification numbers, Dewey numbers and subject headings were still correct. It took me a while!

LC and Dewey Classification numbers (050 and 082 fields)
LC Subject Headings (600, 610, 611, 630, 650 and 651 fields)
LC Authorized forms of authors/editors’ names (100 and 700 fields)

Fig. 3
Also, in these records that require them,
Language and Geographical codes (041 and 043 fields)
LC Authorized forms of Series’ titles (490 and 830 fields)
Corporate or Conference main entry (110 and 111 fields)
Title variants (130, 240 and 246 fields)

These collections are updated every month when new titles are published, because I add new improved MARC records for the new titles to the collections. These manually created records are not yet available through our platform, Cambridge Core. We are working on having them on our platform soon, so that customers can download them and upload them in their library systems without any delays. In the meantime, if customers are not happy with our existing automatic eBook MARC records, they can write to me or their sales representative and we will send them these improved records.

A special project: CLC eBook MARC records

The Cambridge Library Collection (CLC) (Fig. 4) is a beautiful collection of books from Cambridge University Library, with titles not only in English but also in French, German, Italian, Latin and other languages. These titles were all originally published between the 16th and the 20th century, so they are all out of copyright, and they have been brought back to life for the use of students and researchers. The collection was born from a joint project between Cambridge University Press and the University of Cambridge. It’s a collection of significant historical interest and with which we have achieved amazing results in terms of use and sales. But we know that we could achieve better results if the eBook MARC records supplied for this collection were improved.

Fig. 4
In this image (Fig. 5) you can see the structure of these eBooks. There is a new title page, a new title page verso, containing the new and the original print publication date, and a reprint of the original title page.

Unfortunately, our system doesn’t have the data about the original publisher and place of publication, so our records (Fig. 6) look like the example on the left: with “Place of publication not identified” and “Publisher not identified” in the 264 field (elements that are of no help at all for users in their searches); also there are no subject headings; no classification numbers and no other useful access points. On the right there is the same record but upgraded with all the correct access points and this is the kind of record we want to supply for this collection. The process of building these improved records will take me a bit longer than it did for the Cambridge University Press ebooks, because we don’t have LC print records for this collection. So I need to copy (after checking that everything is still correct) what I find in the LC records for the original publication. However, I will still need to check each eBook because many of these titles are two or more titles put together for the first time, so they have never been catalogued together before and it will be necessary to catalogue them from scratch. I aim to complete this project by mid-2017. I hope you can wait that long!
Special project 2: OPL MARC records

My second project is the MARC records for the MRS Online Proceedings Library (OPL) (Fig.7). This is one of the most important resources on the Cambridge Core platform. This resource was originally published as a monograph collection and many customers have requested MARC records that would describe the monograph collection but with links to the online resource. Since none of the 1,565 records were created before within our system, I had to start from the beginning, creating records with MarcEdit and using the print versions that I was able to find on the Library of Congress catalogue and other sources. There were about one hundred titles available in the Cambridge University Press archive and I have catalogued these as part of the process, but finding some of the other records is proving to be more difficult.
As you can see from the image (Fig. 8), we will supply the completed data as RDA records with all the access points necessary. Each record will have a unique number and link to the matching online issue. Again I aim to complete this project by mid-2017.

Cambridge Core: eBook MARC records and KBART lists

This (Fig. 9) is our new platform Cambridge Core. It is beautiful, isn’t it? With a brand new design, it contains all our Academic content, ebooks and journals. Once administrators are set up, they can login and access their institutional account.
In their account they can select the option ‘Bibliographic data’ and then download eBook MARC records, OCLC control numbers and KBART lists (Fig. 10).

For eBook MARC records and KBART lists there are three different options: records by organisation, by order and by identifier (ISBN/ISSN). For third parties and vendors there is a different page (Fig. 11). Without logging in, they can select the ‘Services for Librarians’ on the Cambridge Core homepage and from there download eBook MARC records and KBART lists by collection.
To recap:

- The automated eBook MARC records have been improved at CUP
- There are also manually enhanced eBook MARC records available and customers can get these records from the Library Data Analyst or from their sales representative.
- Two special projects: the CLC and OPL MARC records
- The Cambridge University Press new platform, Cambridge Core

On a personal note, this is me (Fig. 12) with my manager Tristan Collier and our director Jenny Mathias, when I received a very special award from the Press, the Quality Award. It was given to me because the quality of the enhanced eBook MARC records has been greatly appreciated by our customers. Along with the certificate I received this beautiful box of cupcakes! I’ve included this to show that the Press is committed in improving the metadata supplied to customers and third parties.

So now this is a thank you from me! It has been a pleasure to give a presentation at the 2016 CIG Conference and to be able to publish this article on the 2016 December issue of Catalogue and Index. At the bottom of this page you will find my email address. Please write to me if you have any questions, if you want MARC records (obviously) and if you have any suggestions to further develop my projects. I am always happy to hear from fellow cataloguers!

Email: claspada@cambridge.org
BDS is a cataloguing agency, creating bibliographic records to library standards. We have built our database over the past twenty-one years using a data entry system that has received regular upgrades to keep abreast of both technological and bibliographical developments. With its success based upon the quality of its data, BDS always strives to use the most efficient and effective methods for data entry and the recent developments around RDA and linked data persuaded the BDS development team that it was time for a replacement. Our long-standing and highly successful system has its roots in DOS, and uses a programming language which is increasingly difficult to support. It requires training in MARC21 and internal processes, as well as in the standards that determine the content of the bibliographic record. Here, I shall describe the process of designing new software for a cataloguing and associated data entry system which, after a staff competition to name it, is called “Aura”.

BDS implemented Resource Description and Access as the cataloguing standard in January 2013. RDA is based on the conceptual model FRBR, using concepts of works, expressions and manifestations. A work is realised through its expression, which is embodied in the manifestation. It became obvious to us, and the rest of the cataloguing world, that RDA does not fit well in MARC. In some ways this made implementation easier, as it meant smaller changes in practice, but MARC’s flat-file structure doesn’t allow us to fully realise the relationships involved. It is manifestations which we catalogue, and to date we’ve created an individual, isolated record for each manifestation.

The FRBR model underlying RDA, the advent of linked library data, and the idea of a ‘master’ record that we’d had for a while, all combined to suggest a relational database structure for the new data entry system. A relational database holds data in tables which have relationships to one another. Note the use of ‘data’ rather than ‘records’, which is a significant change of mind-set. The ‘cluster’ of associated data for a manifestation could be called a record, but we’re going to have to find ways to think differently about bibliographic information.
The relational database is powerful, because if we change data in one table, all connected data is also affected. For example, when the first manifestation of Phil Vickery’s Gluten-free cooking was catalogued in 2009, there was no specific Dewey number for gluten-free cooking. By the time the 2016 updated edition was published, we did have a specific Dewey number for gluten-free cooking. By updating the Dewey of the work, we would update the Dewey for every manifestation of that work.

It’s also possible for BDS to have records for multiple manifestations of a work, but only catalogue one of them book-in-hand. The remaining manifestation records are currently based on the publisher’s information at the CIP stage. With a joined up database, any confirmed details which affect the work and expression for the manifestation we have in hand will then also positively affect the data for the other manifestations.

Whilst many manifestations are part of a network of connected resources, other publications stand alone and so there’s no existing data to connect to. This has to be borne in mind to make sure the new system is still efficient when all of the data has to be created from scratch.

So, we decided that the new database and interface should:

- use the FRBR model
- be able to produce MARC records
- bear in mind developments for post-MARC bibliographic data
- support rather than hinder productivity
- be able to import data, both MARC data from the existing database and ONIX data from publishers.

The data had to be as granular as possible, and the interface had to have as much of our cataloguing expertise built in as we could write logic statements for. The work is far from over, but there have been three main areas of effort: the model for the database, the ‘translations’ to import and export MARC records, and the business rules which support the creation of good data. I’ll outline the kinds of issues involved with each of those.

We began by listing all the bits of information we wanted to include, and deciding which table they belonged to. In addition to the RDA elements, we needed to capture additional information that the MARC record would need, and BDS requirements such as trade edition statements. An easy, early decision was to use as many identifiers as possible. Each entity having a unique identifier is inherent in the relational database structure, but we also want to store persistent global identifiers like ISNIs, and NACO control numbers. This supports future linked data developments, whether using Bibframe or other standards.
The RDA implementation of the FRBR model is only dealing with descriptive cataloguing, and not with classification and indexing. There were problems to solve where that model interfaced with other knowledge organisation systems. BDS uses Library of Congress Subject Headings (LCSH) for subject indexing, and we saw the subject of a resource as being at the work level, but we couldn’t simply assign all LCSH to the work level in the model. This is because LCSH include form headings such as “Talking books” and “Large type books” which pertain to expression and manifestation levels respectively. Our solution was to allow expression and manifestation level LCSH.

We also realised we had an issue with aggregate works. Aggregate works are anything that can be split into separate works, so a book containing articles by different authors is the obvious example. There are many other examples, though: omnibuses of novels, continuing resources, and resources which include the same work in different expressions. When one thinks about each work included in the aggregated work, and all the potential relationships between those works, the model gets complicated. The principal problem we had to solve was self-referencing works, i.e. a publication which both is and is about a particular work, such as the text of Hamlet plus study notes, or a Bible which has commentary as well as the Bible text. Our pragmatic solution is to create works where we need them, and avoid self-referencing.

After resolving modelling issues, the next step was to tackle the interface with MARC. Importing data that’s less granular than your database is problematic; exporting requires a high level of complexity, but it does fall to logic - so long as the database is granular enough.

The 245 Title field and the 300 Physical description fields are both examples of difficult areas in MARC, because the subfields contain more than one kind of data. A human looking at a 245 field can determine which full stops are there for abbreviations, and which are there for ISBD punctuation, but a machine cannot:

<table>
<thead>
<tr>
<th>Incoming MARC</th>
<th>Aura database</th>
<th>Outgoing MARC</th>
</tr>
</thead>
<tbody>
<tr>
<td>245 04 $a The history of the world. $n Volume 2, $p Kingdoms and empires</td>
<td>Title=The history of the world Filing characters=4 Part number=Volume 2 Part title= Kingdoms and empires</td>
<td>245 04 $a The history of the world. $n Volume 2, $p Kingdoms and empires.</td>
</tr>
<tr>
<td>Title=The history of the world Filing characters=4 Statement of responsibility=edited by J. Smith Part number=Volume 1 Part title=The first civilizations</td>
<td>245 04 $a The history of the world / $c edited by J. Smith. Volume 1, The first civilizations.</td>
<td></td>
</tr>
</tbody>
</table>

In the 300 field, the same fields are used differently for different kinds of material:

```
300 $a 9 CDs : $b digital, stereo
300 $a 456 pages : $b illustrations (black and white) ; $c 24 cm
```

Both of these fields require a lot of complicated programming when importing, and a lot of information captured specifically to build the MARC for export, such as detailed labelling of the Title statement. An icon produces a MARC view of the record in “Aura”, allowing users to check the output.

Last but not least is the business rules, which in this context means the kind of rule that says “If you’ve identified this resource as having the Main content type of ‘text’, then you cannot have data in the field for Expression duration, because that only applies to Spoken word, Performed Music, or Moving images.” As you can imagine there are a lot of logic statements which could be deployed, and we’re still adding to them.
There are some basic defaults, which 9 times out of 10 will save a few keystrokes. For example, the language of the expression defaults to English. Then there are conditional defaults, for example if the work form chosen is Music then the main content type default changes from text to notated music.

The final component to the business rules is error checking – a work has to have a title, and other core pieces of information, so the system won’t let you save it unless the core requirements are met. If the physical format chosen is incompatible with other data elements, the system can flag that up to the user.

The user interface still needs a lot of work and testing, but a significant change for us is the use of a tabbed layout. Since we’re not constrained by MARC order, a tabbed layout optimises the visibility of data elements and their accessibility for entering and editing by the user. There’s a Work, Expression, and Manifestation tab, and each of those has tabs for specific aspects of the data.

There’s a search box, and double clicking on a result brings up the relevant information in the main screen, and a WEM tree in the bottom left.

If I search for an ISBN, and it’s not in the database, the system looks for it in the old database and the publisher’s ONIX database. When a record is found, the user then contributes to how it’s imported. The system attempts to match to works and expressions already in the “Aura” database, and the user can accept the suggested matching work, or choose a different one, or demand a new work is created. If an existing work is selected, the system goes on to look for matching expressions. If you decide it’s a new work, then obviously it will need to be a new expression too. If matching works/expressions have been found, then we have far less to do than creating the record from scratch.

A significant training issue for this new system will be to ensure users realise the effect that changing information in just one field can have. Depending on where a change is made, a cascade of differences could result. There will therefore be different levels of user status, with only the most skilled and trusted being able to make high-impact changes. There are also decisions to make about which field changes are worth triggering a “resend record” status for. For example, updating the Dewey of a work is useful to virtual retrieval, but is not going to change the spine labels on items on library shelves.

If cataloguing becomes more about identifying entities and their relationships and we move from text based bibliographic records, to use as many identifiers and discrete pieces of data as we can, I think cataloguing will be a more intellectually satisfying process. It will also facilitate better search and retrieval for users. Creating “Aura” continues to be an interesting and exciting project where BDS, in line with its ongoing philosophy of quality and accuracy dedicated to accommodating the future, is leading the way in its field. Now, we are looking forward to getting to the testing stage and further refining “Aura”.
I would like to begin my lightning talk today by thanking the organizers of this conference for the opportunity to describe what the IRIS Consortium of Florentine area member libraries are doing regarding RDA records coming into our union catalog. There is a wonderful reflexive Italian verb, *barcamenarsi* which, as with many words in that rich language, has various shades of meaning. I am using it here to mean how we are managing, coping and acting cautiously in this transitional period.

First a bit of background: The IRIS Consortium, founded in 1993, is an association of Florentine area art history and humanities libraries consisting today of six members. The founding members are: the Berenson Library (Villa I Tatti - The Harvard University Center for Italian Renaissance Studies), the Library of the Dutch University Institute for Art History (the presence in Florence of Utrecht University), the libraries of the Gallerie degli Uffizi which include the main library of the Uffizi, the Prints and Drawings Library, and the six specialized libraries at Palazzo Pitti, and the Library of the Fondazione di Studi di Storia dell’Arte Roberto Longhi (1993-2015). The consortium has expanded with the arrival of the Library of the Istituto Nazionale di Studi sul Rinascimento (1997), the Biblioteca «Ugo Proacci» of the Opificio delle Pietre Dure (focused on the literature regarding the conservation and restoration of art objects) (1999) and, finally, the Leonardiana Library at Vinci (2007), the library of record for a vast bibliography on all aspects of Leonardo studies.

We are clients of ExLibris: our ILS is ALEPH Version 22. Our catalogue consists of over 352,000 bibliographic records plus a file of some 143,000 authority records mainly from the Library of Congress. Our cataloguers use AACR2 in English and its translation in Italian. We are a bilingual catalogue: the American and Dutch partners catalogue in English, the Italians partners in Italian. For subject analysis the American and Dutch partners use the Library of Congress Subject Headings (LCSH) while the Italian cataloguers provide subject access using terms from the Soggettario BNCF, the thesaurus maintained by Biblioteca Nazionale Centrale di Firenze. Our sources of records in RDA are: monthly file loads from the Berenson Library cataloguers, and LC records available to all our cataloguers via the Z39.50 protocol. As of the writing of this paper the number of RDA records in our union catalogue is 4,556. Our presence in WorldCat via OCLC’s initiative “Art Discovery Group Catalogue” is a point of pride for us: ca. 325,000 of our records point the end-user to the IRIS catalogue.

As the Cataloguing Specialist for the consortium, I monitor the AUTOCAT and RDA-L listservs whose lively discussions and e-forum offerings have permitted me to have a decent, if general, overview of the cataloguing situation as many libraries make the transition from AACR2 to RDA. As time passed, however, my colleagues and I realized that the moment had arrived to give both our cataloguers and library staff with reference duties an overview of what was happening in the RDA world, and what our decisions and positions would be with regard to that situation. Basically, the what, the why, and the how we were positioning ourselves in this evolving environment. A mini-course was organized inviting all our staff to attend. One colleague gave a brief overview of RDA and FRBR, the Head Cataloguer at the Berenson Library analyzed a typical RDA record present in the IRIS catalogue, and the Head of Cataloguing at Casalini Libri explained the changes to be seen in RDA authority records. I concluded the event with a list of dos and don'ts for our catalogers when dealing with RDA records, and “painless” suggestions for making our records more user-friendly.

This brings me to the reason for my use of RDA-Lite in the title of this presentation. Our decision is to not take the “full plunge” into using RDA as a cataloguing code for original cataloguing, but to successfully accommodate incoming RDA records, assuring a co-existence with our AACR2 “legacy” records. As Gordon Dunsire wrote in his conclusion to his article in a the recent issue of JLIS: the “… economic constraints faced by the entire cultural heritage sector” was, and is, particularly true of the situation in Italy. Libraries large and small, public, private, academic, those depending on funds from the state or other sources, have all been hit hard.
Our reasons for not adopting RDA will sound very familiar to those of you who have had to evaluate your local cataloguing situation. For the IRIS Consortium the annual subscription with multiple licenses to the RDA Toolkit, the retraining and follow-up of staff who more often than not have multiple duties not involving cataloguing or, indeed, are staff consisting of one person, were reasons too imposing to ignore. The necessary disruption of cataloguing “productivity” in collections where massive amounts of material remain to be put online was an additional reason too imposing to be ignored. Taken together, the decision in favour of our transition to RDA could not be justified to our board of directors. However to this rather negative list I must add a positive: that our holdings visible in WorldCat via the Art Discovery Group Catalogue were guaranteed a continuity by OCLC’s realistic and pragmatic decision to continue accepting AACR2 records.

The arrival of records in RDA into our catalogue has caused us to take a fresh look at our cataloguing practices to see what we could be doing to make our records more user-friendly.

Here are some of our cataloguing decisions:

- Our cataloguers have been instructed to avoid the use of abbreviations not only in the 300 field, but also in the 5XX note fields. This is particularly important for us because notes may be added to any record in either English or Italian.
- Also to be avoided if at all possible is the use of “S.l.: s.n” in Field 260, instead making more use of the internet to verify information regarding place and publisher or the entity responsible for the publication.
- We will continue to use the 260 field, but will add the 264 field when the copyright date differs from the date of publication. (The 264 field displays the label “Copyright Notices” so its appearance should not cause confusion to the end-user.)
- Our cataloguers are encouraged to make use of the 520 field, especially when the title of the work and/or the “creative” graphics on the title page in hand seem designed more to attract attention than to describe actual content. Those of you who catalogue art-related publications, especially art exhibitions, will know what I mean!

What are we doing with regard to RDA records coming into our catalogue from the Berenson Library or from the Library of Congress?

- We will not edit RDA records back to AACR2.
- We will not delete RDA fields 336, 337 and 338. They do not display in our OPAC but may be useful in a future which we cannot for now predict.
- We will not delete the sub-field relationship designators (sub-field “e”) from the access points where they appear.
- However, for our original cataloguing, we will not be adding sub-field “e” to the access points, but we will continue to “justify” the presence of these access points - when necessary - with brief notes.

Regarding authority records in RDA

We have noted with much appreciation the richness in detail of authority records created according to RDA guidelines. The “downside” is, of course, the time necessarily needed to bring together and provide this level of detail. The result is that one of our partner libraries has decided to cease creating and contributing authority records for its original cataloguing when a needed authority record is lacking. We wonder if there isn't some “middle way” to proceed to avoid decisions, however necessary, which penalize the goals of clear identification and uniform access to materials in library catalogues.
Regarding the situation in Italy for cataloging in RDA

A review paper for EURIG (the European RDA Interest Group) by Alan Danskin and Katharine Gryspeerdt entitled “Changing the Rules: RDA and Cataloguing in Europe” stated that as of the time of writing several translations of RDA were underway. The Italian translation was among those listed, and was added to the RDA Toolkit in 2015. The translation, eliminating a linguistic barrier, owes its existence to the concerted effort and dedication of a group of nationally appointed Italian cataloguing experts from major public and private libraries, university libraries, bibliographic agencies and the Vatican Apostolic Library. I would like to note here the impressive contributions to the translation by Professor Mauro Guerrini of the University of Florence and Professor Carlo Bianchini, University of Pavia, and their team of research assistants.

Professors Guerrini and Bianchini have also fostered promotion of RDA in Italy for a number of years by publishing widely, organizing courses, lectures, and seminars in Florence and elsewhere in Italy featuring prominent figures in the development of RDA such as Barbara Tillett and Gordon Dunsire. The title of an article by Prof. Guerrini nicely sums what their efforts are aiming for: “RDA in Italian: an Opportunity to Join the International Context”. A recent issue of JLIS, the online Italian Journal of Library and Information Science, was dedicated to this very topic. Articles included overviews and critical studies of RDA in Italian translation. And space was accorded to the “opposition”: a contribution by Michael Gorman provocatively entitled “RDA: the Emperor’s New Code”.

So, given the interest, promotion and dissemination of RDA who is actually currently cataloguing using RDA as a content standard? The answer, as of the writing of this paper, is: our consortium partner, the Berenson Library at Villa I Tatti, the Harvard University Center for Italian Renaissance Studies, and Casalini Libri in Florence, the highly-regarded vendors of a suite of library services, which include RDA bibliographic and authority records, for their English language clients. As time goes on it will be interesting to see who, how and when other entities make the transition. In the meantime I would like to offer a thought, perhaps a possibility in some form.

The Dutch “solution” described in the EURIG paper I referred to a moment ago intrigued me. In the place of a full RDA translation in Dutch a document aimed at the Dutch cataloguing community proposes to provide a “work flow or interpretation layer” … “sufficient to meet the needs of most cataloguers, most of the time” … Could perhaps an Italian language version of this approach be the useful helping hand for the Italian library cataloging community? Guerrini and Bianchini and the Italian cataloguing experts involved have already done the heavy work: can they, working with the Italian cataloguing experts, perhaps take it forward with something like a “Dutch solution” to the benefit of all?

Thank you, and I welcome, at this point, your questions and observations.
Background

CILIP has asked its member networks and Special Interest Groups to consider equality and diversity in their business plans. The Equality and Diversity Strategy Project looked at CIG’s current membership profile, identified potential barriers to participation, and assisted in developing CIG’s strategy to address equality and diversity. This project was conducted by participants on the CILIP Leadership programme,¹ on behalf of the Cataloguing and Indexing Group.

The project surveyed CIG members and benchmarked this against the wider information profession. We also identified possible barriers to participation in member network activities, and created a toolkit to support other member networks in analysing their own membership and developing their equality and diversity strategy.

Literature Review

Our review of the published literature concluded that while there was a great deal of work focused on diversity and inclusion within professions as a whole, there was not much that tackled the activities of the professional bodies themselves. However, these two issues are closely linked.

Information-related professional bodies such as the Young Adult Library Services Association have been looking at how to increase the diversity of the profession (Craig, 2010). Similarly the Association for Library and Information Science Education has made suggestions for improving cultural awareness in the information studies curriculum (Lee et al., 2015). Newman (2016) points out that the American Library Association’s “top tech trends” panels are consistently male-dominated, despite the organisation’s membership being 81% female, and recommends conference organisers be more pro-active in seeking diverse panels.

Looking outside the information profession, useful advice can also be found in the publications of The Minerals, Metals & Materials Society, who now include improving diversity and inclusion in their strategic plan (Robinson, 2015). In 2014 the society held a summit on diversity, producing both a toolkit (The Minerals, Materials & Metals Society, 2015a) and a lengthy report (The Minerals, Materials & Metals Society, 2015b). The Society for Conservation Biology also provides practical suggestions for increasing diversity in their profession (Foster et al., 2014).

A lengthy report commissioned by the ASAE Foundation (Leiter, Solebello and Tschirhart 2011) makes suggestions for increasing diversity and inclusion in membership associations, including a review of projects that did and didn’t work (p.51). An article in the Journal of the Academy of Nutrition and Dietetics (Stein, 2013) reviews the diversity projects of various professional bodies. Many of these professional bodies declined to share measurable outcomes of their projects, and the Stein paper concludes that whilst there isn’t a single solution for improving diversity, various options do exist.

¹. The CILIP Leadership Programme was designed to help mid-career professionals wanting to develop their skills and ran from July 2015 to July 2016.
Benchmarking CIG membership

A survey of CIG members was launched in mid-January 2016 and concluded in mid-March. It was promoted via CIG’s newsletters, Twitter account and blog, and had two sections. One focused on the demographics of respondents, and the other about how they engaged with CIG.

The findings from this survey were looked at alongside two other sets of data: the 2014 Workforce Mapping Survey, commissioned by CILIP and the ARA to map the workforce across the Library, Archives, Records, Information and Knowledge Management Services and related professions in the United Kingdom; and CILIP’s own brief demographic information about CIG members, which had been gathered as part of the CILIP membership process.

At the time of analysis, only certain aspects of the Workforce Mapping Data survey results had been made publically available. One of the aims of the project survey was to compare CIG’s demographics with the wider information profession, so we focused our survey on the publicly available Workforce Mapping Data. This meant that certain themes often cited in diversity and equality research, such as health, disability, family and caring commitments, were deliberately excluded from our survey.

CIG’s own existing membership data comprised of the age and gender of its membership. We had hoped to gather more detailed demographic information about CIG’s membership via the survey, and then compare this to the Workforce Mapping Data project, but the low response to our survey meant that we could not use our data to predict the wider make-up of CIG, and therefore demographic comparisons were limited to age and gender.

Gender
The Workforce Mapping Study identified that the information sector has a 78.1% female make-up. CIG’s demographic data suggests that their membership is 69% female and respondents to our survey of CIG were 72% female. These figures would indicate that the group is attracting more males as a proportion of their membership compared with the profession as a whole.

<table>
<thead>
<tr>
<th>Equality and Diversity project survey</th>
<th>CIG demographics (from CILIP membership application)</th>
<th>Workforce Mapping Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender identity(^2)</td>
<td>72% female</td>
<td>69% female</td>
</tr>
<tr>
<td></td>
<td></td>
<td>78.1% female</td>
</tr>
</tbody>
</table>

Table 1: Gender identity of survey respondents, CIG members and wider workforce.

Age
The workforce mapping survey found that 55.3% of the information and archives workforce is over the age of 45. Unfortunately CILIP’s CIG membership data is not directly comparable as the age ranges don’t match. From CILIP’s data the proportion of CIG members who are over the age of 40 is 78% and 54% of CIG members are over 50.

\(^2\) Gender analysis was further complicated as the CILIP demographics and the Workforce Mapping Study only offered male and female as gender options, but the Equality and Diversity Project included an option for non-binary and prefer not to say.
According to these figures CIG membership is older than the general information workforce. However, of the respondents to the Equality and Diversity survey, only 40% were over the age of 45, hence younger CIG members were over-represented compared to the membership in general.

<table>
<thead>
<tr>
<th></th>
<th>Equality and Diversity project survey</th>
<th>CIG demographics (from CILIP membership application)</th>
<th>Workforce Mapping Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>40% over 45 years old.</td>
<td>78% over 40 years old.</td>
<td>55.3% over 45 years old.</td>
</tr>
</tbody>
</table>

Table 2: Age of survey respondents, CIG members and wider workforce.

**Other indicators of diversity**

CILIP has not been asking its members about their ethnicity, religious identity, or sexual orientation, so this information isn’t available for the wider CIG membership. Survey results about ethnicity, religious identity and sexual orientation are not covered in this paper but were included in the report submitted to CIG, which will be made available to all members.

Some tentative conclusions can be drawn about CIG membership in relation to gender and age. It would appear that overall, in comparison to the wider information workforce, CIG has a higher proportion of male members with approximately 9% more members being male than one would anticipate if the nationwide pattern were followed. With regards to age, taking into account the discrepancy in the way the two figures were measured, it is still possible to deduce that CIG has older members than the information sector in general.

The project group’s survey was based on a self-selected sample, and as with all self-selected samples, the respondents have to be motivated to respond, which may further affect the representativeness of the results. With that in mind, the results collected in relation to identifying barriers may be more valuable.

**Identifying Barriers**

Our survey also looked at barriers that limit member’s participation in CIG offerings. It covered themes such as attendance at events, involvement in e-forums, and access to monthly newsletters. Follow up questions were included to try and capture reasons for limited, or lack of, participation in activities. All respondents were also asked to offer suggestions on how CIG might remove barriers and widen its appeal.

Many respondents expressed a desire to participate in events, but faced obstacles in doing so. The three primary barriers identified were cost, location and timing. A lack of employer engagement was also a significant concern.

Cost is a major obstacle for members. This includes the cost of the event itself, the cost of travel to a free event, or a combination of paying for the event and getting there. From our survey a high proportion of participants, 65%, had been unable to attend a CIG-organised visit, with 24% of those stating that costs and location of visits were the primary reasons. The survey results showed that 47% of participants have attended a course organised by CIG, but of the 41% who have not, 10% stated that the location and costs involved were a barrier to attendance, with another 29% stating they had not been able to attend, despite wanting to.
Funding for training was an issue for many of the respondents, especially when employer funding was not available. There were suggestions that employers are happier funding courses directly related to current work issues, such as RDA, and that CIG may need to work harder at promoting the benefits of other types of CPD to employers.

Location, and specifically the distance to a location, was highlighted as a barrier. Thirty-five per cent of respondents stated that distance and getting time off were significant barriers to participation. This is more than the 24% who indicated cost was a primary barrier, but as the two are often linked, the results shouldn’t be considered in isolation to each other. Some survey respondents reported difficulties in attending events at the locations currently used by CIG, especially those not accessible via public transport.

Timing was the third major barrier to participating in CIG activities, and unlike cost and location, which primarily related to attendance at events, timing also accounted for an inability to participate in online activities. Heavy workloads and small teams were common barriers, with people either unable to get time off due to an inability to arrange cover, or because they simply didn’t have enough time to spare. Many responded that they could only participate in their own time and had to take annual leave to attend events.

Survey respondents gave various suggestions for improving engagement between CIG and its members. These are useful for anyone organising an event (whether or not for CIG) and should be carefully considered. It’s important to remember that making a single change would not help every survey respondent, but reflecting on the feedback and publicly taking on board the suggestions will improve the relationship between CIG and its members. Publicity material for events could highlight that the survey results have been considered, this would show members that their feedback is valued.

We also asked for suggestions on how CIG could better engage with the wider profession and attract potential members. Some respondents were concerned that CIG was not doing sufficient outreach outside of the special interest group, particularly to those who were not working in traditional libraries or were not members of CILIP.

**Toolkit**

We also produced a toolkit that could be used by other member networks to inform their strategy on equality and diversity. The toolkit was broken down into three sections. The first section focuses on analysing members, drawing on our experience of running the survey for CIG. This section gives detailed advice for gathering information, and the different types of approaches groups might want to take.

The second section focuses on identifying barriers, and includes tips on identifying less obvious barriers. We felt that given the survey responses, it would be the ‘hidden barriers’ that would be likely to pose more of a problem to CILIP groups than the more visible and clearly legislated ones.

The third section of the toolkit focused on developing strategies. It was not designed to offer solutions, but instead to highlight some of the options which might be available. It was important that we created something that was fluid enough to adapt to individual groups, but not so vague as to be of no use. This is because any equality and diversity strategy must be specific to the group and take in the unique nature of their membership. It must also be achievable and practical, but flexible enough to address changing needs. It’s important to not assume that all members identifying as a particular group will have the same needs, or want to be treated the same.

Finally we also provided a reference list of sources of information covering aspects such as Planning accessible information from the Sensory Trust, to Uncovering hidden impairments from the Hidden Impairment National Group.
Conclusions

This project identified a number of steps that CIG can take to address issues surrounding equality and diversity. Taking these steps will hopefully increase engagement with CIG’s membership offering. The recommendations of this project include measures of differing intensity, as well as short- and long-term options. This will enable CIG to implement some measures immediately, whilst some will require a more considered and strategic approach. The project also produced a toolkit that can be adapted and used by other CILIP Special Interest Groups.

At the conclusion of this project in July 2016 both the final report and the finished toolkit were submitted to the CIG committee who will be able to make them available to any CIG members who wish to see it. The Equality and Diversity in Member Networks Project team enjoyed working on what was a very interesting topic, and we are grateful to CIG for commissioning it, and to those members who helped promote, and responded to, the survey. We look forward to seeing where CIG decides to take the project in the future.

References


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: Helen Garner, (h.j.garner@shu.ac.uk) and Karen Pierce (PierceKF@Cardiff.ac.uk)

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

ISSN 0008-7629

CIG website: http://www.cilip.org.uk/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