ISSN 2399-9667

Catalogue and Index

Periodical of the Metadata & Discovery Group,
a Special Interest Group of CILIP, the Library and Information Association

Editorial

June 2024, Issue 208

Welcome to C&I 208,

This issue presents papers from, and about, the 2023 MDG biennial conference. The theme was “ReDiscovery”, and the event took place on Wednesday 6th and Thursday 7th September 2023 at IET Birmingham, followed by the RDA day on Friday 8th.

The reach of the conference, as demonstrated in the papers in this issue, is global and significant. An example of this is the update by various members of US institutions on their work on DCRMR, the core rules amended in response to RDA for SCA use, with a global reach in terms of use and impact. The importance of the conference as an occasion for metadata practitioners of widely ranging levels of expertise and areas of interest to share knowledge in a friendly and supportive environment is perhaps best demonstrated by the final two articles: Keynote speaker (and Alan Jeffreys award recipient) Alan Danskin shares his memories of working at the British Library for 36 years, whilst bursary winner Jaycie Carter writes about her experiences attending her first MDG conference.

This issue is the first to be published on our new platform, hosted by Edinburgh Diamond. It has been a long journey getting here, but we hope that the improvements over our old solution (and the ability to finally practice what we preach, metadata-wise!) will make the wait worthwhile.

Philip Keates: p.keates@kingston.ac.uk

Martin Kelleher: martind@liverpool.ac.uk

Contents

MarcEdit task lists and vendor-supplied metadata: revisiting the supplier-vendor relationship at the University of Leeds

Ceilan Hunter-Green, Metadata and Discovery Coordinator, University of Leeds Libraries

Like all other academic libraries, the University of Leeds needed to hugely increase our electronic provision after the first Covid lockdown in March 2020. Most of the pressure was on providing e-textbooks and e-books, but there were also many modules that suddenly needed access to streaming video. One of these streaming video suppliers was the British Film Institute, who launched a new Institutional Subscription offer during lockdown, but could not provide records. This provided an opportunity for Leeds' Metadata and Discovery team to develop a process of automated record creation using MarcEdit templates and task lists and provide those records to the subscribing community at no increased cost, whilst receiving a discount on our own subscription.

British Film Institute subscription

The BFI streaming video collection is strong in areas that our other institutional suppliers aren't, and their films were particularly needed on Japanese and history of film modules where foundational material is difficult to come by otherwise. There were some challenges, though, as their institutional subscription was new at the time and the streaming video service's metadata format was very different to what we were used to as librarians. They also have a comparatively small collection with around 600 active titles at any given time, which is updated very frequently compared to the more static collections we were used to. Most significantly to us, they didn't offer MARC records, which meant that we spent a lot of time on maintaining local records for their films so that we could link reading lists to individual catalogue pages rather than linking to the BFI Player site for students to repeat their search. But that lack of MARC records became an opportunity for us to offer something back to them as vendors.

For the first year of our subscription we, like all of BFI's other subscribing institutions, received metadata for the streaming video collection's films in the form of a very basic spreadsheet, originally 10 columns and upgraded to 12 columns in February 2022. We used MarcEdit's Delimited Text Translator to create bare-bones records, and then as an Ex Libris library running Alma as our LMS we used Alma's import profiles to run normalization rules on those basic records to get them into our system in semi-decent MARC shape.

The provided metadata covered the film's title, internal BFI ID number, access start and end dates, country of origin, release year, two Genre terms and two Director names, though usually the films just had one director. The remaining metadata that we needed wasn't provided—so it was copied from the BFI Player streaming platform field by field into Alma's metadata editor with each monthly update. The average monthly addition was 22 films, but in practice this meant some months had five additions and some months had fifty, so it was an unpredictable demand on our staffing resource.

Renegotiation and enhanced metadata provision

This time-consuming and inefficient record production continued for a year, until colleagues in the Acquisitions team were preparing to go into negotiations with BFI for a new year's subscription. Previously BFI had asked about the process of us providing the records but it didn't feel practical when the process was so time-consuming. But as we went into that second year of subscription, with more familiarity with the metadata from BFI Player, more experience with MarcEdit and more confidence in the value and usefulness of our records to the community, we thought we would be able to take another look at the process of record creation.

At the University of Leeds, purchasing is handled by a separate Acquisitions and Reading Lists team, and my Metadata and Discovery team were responsible for handling the access and the discoverability of these resources once we had them. Acquisitions colleagues studied the average price for MARC records that we paid to other subscribers and the amount of time we were spending on the process of record creation, and were successful in negotiating a discount. At that point, we just had to develop a workflow to create the records in a much more efficient way than we had been; we needed much more (and better quality) metadata to start with, and the resulting records had to be an even higher standard if they were going to be shared with other institutions.

The team at BFI worked with their systems team to extract all relevant metadata from their streaming platform in the form of a 56-column .csv file, a huge improvement on the previous spreadsheet. We now have metadata for all cast members in separate columns, a unique system identifier from BFI, runtime, rating, and everything else we need which previously had to be copied from the BFI Player site to fill out our records. We also started to receive some additional metadata like original language titles in addition to English titles, which isn't on the public BFI Player site, and a direct URL instead of the general landing page URL we added before.

MarcEdit transformation – import template

To translate this 56-column spreadsheet into a MARC record, we start with the MarcEdit import template, which brings that mammoth spreadsheet into MarcEdit (Figure 1) through the Delimited Text Translator. It mostly imports one column per field with a few exceptions where data from a single column is added to two fields, such as the Cast column data added both to a concatenated 511 and to individual 700 fields, and the Year of release data added both to the 501 and the 046. Some fields are populated as placeholders, like the 501 for the year of release which will be moved to a 500 later, and the directors are added to a 701 until the addition of a relator term, when they're moved to a 700 as well. Most of the data at this point is pretty blunt.

```
=LDR  00000ngm a2200000la 4500
=008  230831s9999||||xx\|||\|\\\\\\|\\\|und||
=024  7\$a9afb773d-ce2e-59dd-b516-b13c84dada79
=046  1\$k1953
=130  0\$aTōkyō Monogatari
=245  00$aTokyo Story.
=300  \\$a136
=500  \\$aJapan.
=501  \\$a1953
=506  1\$a02/09/2025;@.
=511  1\$aChishu Ryu,Chieko Higashiyama,Setsuko Hara,
=520  \\$aA constant fixture in critics' polls, Yasujirō Ozu's most enduring masterpiece, Tokyo Story, is a
beautifully nuanced exploration of filial duty, expectation and regret. From the simple tale of an elderly
husband and wife's visit to Tokyo to see their grown-up children, Ozu draws a compelling contrast
between the measured dignity of age and the hurried insensitivity of a younger generation.
=521  8\$aU
=532  1\$aNo
=540  \\$aNo
=546  \\$aJapanese
=650  \4$aAgeing
=650  \4$aFamily
=653  \6$aDrama
=700  1\$aChieko Higashiyama
=700  1\$aChishu Ryu
=700  1\$aSetsuko Hara
=701  1\$aYasujirō Ozu
=856  40$uhttp://player.bfi.org.uk/subscription/film/watch-tokyo-story-1953-online
```

MarcEdit transformation – task lists

MarcEdit's Task List functions let users join together lists of tasks in order to always run them in the same order and the same way for every set of records. Instead of having to create the tasks from scratch and type out the replacements and regular expressions to use every time, the program will remember the tasks and the order and run them all at once, and it's also possible to share these task lists (like the templates) between users in order for all colleagues to perform the same tasks in the same way across the team. It's incredibly useful!

For our BFI data we run two task lists. This first set of tasks adds the Library of Congress fields for Short or Feature film to a new 655 field depending on the run time in the 300. It's a bit silly that it has to be a separate task list, but when those tasks were integrated into the second list, the second list started to delete the 856 field from every 13th record, so we've given up for the time being.

This first task list has a simple effect but uses a kind of logical puzzle to get there.

First it copies the text of the 300$a into a new 655 if the number in the 300 is over 40, which is the Library of Congress run time threshold for short films. Then it replaces the text of *any* 655 with Feature Film, because that new field is the only 655 in the record at this point. It then adds a new 655 field for Short films to all of the records, adds the necessary $2 to the Feature films 655, then deletes the Short film field (replaces it with nothing) if a Feature films field is also present in the record. It goes around the houses, but gets there in the end.

Tasks:

```
COPY    300     655     false   (\$a)([4-9][0-9]|\d{3,})        $1$2    true
REPLACE         (=655)(.{6})(\d*)       $1  \7$aFeature films.          2               0       False
ADD     655     \7$aShort films.$2lcgft         100
REPLACE         =655  \7$aFeature films.        =655  \7$aFeature films.$2lcgft         0               0       False
REPLACE         =655  \7$aShort films.$2lcgft           0       =655  \7$aFeature films.$2lcgft         1       False
REMOVEBLANKLINES
```

Figure 2. Task List 1, 655 for Short and Feature film

Here's what our record looks like after running that first task list (Figure 2). The run time in the 300 field is 136 minutes, so this record needs a Feature Films field. The rest of the record is the same as before, only a that new 655 has been added.

```
=LDR  00000ngm a2200000la 4500
=008  230831s9999||||xx\|||\|\\\\\|\\\|und||
=024  7\$a9afb773d-ce2e-59dd-b516-b13c84dada79
=046  1\$k1953
=130  0\$aTōkyō Monogatari
=245  00$aTokyo Story.
=300  \\$a136
=500  \\$aJapan.
=501  \\$a1953
=506  1\$a02/09/2025;@.
=511  1\$aChishu Ryu,Chieko Higashiyama,Setsuko Hara,
=520  \\$aA constant fixture in critics' polls, Yasujirō Ozu's most enduring masterpiece, Tokyo Story, is a
beautifully nuanced exploration of filial duty, expectation and regret. From the simple tale of an elderly
husband and wife's visit to Tokyo to see their grown-up children, Ozu draws a compelling contrast
between the measured dignity of age and the hurried insensitivity of a younger generation.
=521  8\$aU
=532  1\$aNo
=540  \\$aNo
=546  \\$aJapanese
=650  \4$aAgeing
=650  \4$aFamily
=653  \6$aDrama
=655  \7$aFeature films.$2lcgft
=700  1\$aChieko Higashiyama
=700  1\$aChishu Ryu
=700  1\$aSetsuko Hara
=701  1\$aYasujirō Ozu
=856  40$uhttp://player.bfi.org.uk/subscription/film/watch-tokyo-story-1953-online
```

Figure 3. The original record with a newly added 655 field for Feature films.

The second, more robust task list is 179 tasks long and handles all of the rest of the transformation from essentially a spreadsheet in MarcEdit form into proper records. It does a few different types of tasks from really basic ones to more nimble ones that cross-reference multiple fields.

Those basic tasks include standard field additions to every record in the file, which aren't conditional on the existence or content of any other fields, and which don't involve altering the order or the content of the field text. So with these tasks we add things like the 006 and 007, the 336, 337, and 338, 347 for the video file format, a 264 \2 for BFI's distribution, a 506 field to say that access is limited to within the UK, and a 588 field to indicate that we've constructed these records from vendor-supplied metadata.

The task list then does some indelicate, brute-force amendments. Library of Congress wants the Country of Production to be present in a 257 field, and those countries should have standardized names according to the Name Authority File source (some of which are not how BFI provides them), so the task list converts them into the correct format. For example BFI might say the country of origin for a particular film was the USSR, but the NAF name is Soviet Union, so this set of tasks (Figure 4) standardizes those in the 257 field.

```
EDITFIELD257USA0United States
EDITFIELD257USSR0Soviet Union
EDITFIELD257Federal Republic of Germany0Germany (West)
EDITFIELD257Russian Federation0Russia (Federation)
EDITFIELD257People's Republic of China0China
EDITFIELD257Republic of Korea0Korea (South)
EDITFIELD257State of Palestine0Palestine
EDITFIELD257Socialist Republic of Vietnam0Vietnam
EDITFIELD257Burkina Faso (Upper Volta)0Burkina Faso
EDITFIELD257Canada (Quebec)0Canada
```

Figure 4. Sample of the tasks to convert BFI country names to LoC format.

Another brute-force task handles the language code in the 008 (Figure 5). When we first received the bulk of BFI's records in May, we pulled out every language represented in their collection regardless of whether the film was active on the streaming platform, and added these all into the task list, cross-referenced with the MARC language codes.

```
REPLACE||und||||alb||0=546  \\$aAlbanian1False
REPLACE||und||||amh||0=546  \\$aAmharic1False
REPLACE||und||||ara||0=546  \\$aArabic1False
REPLACE||und||||arm||0=546  \\$aArmenian1False
REPLACE||und||||bam||0=546  \\$aBambara1False
REPLACE||und||||ben||0=546  \\$aBengali1False
REPLACE||und||||cat||0=546  \\$aCatalan1False
REPLACE||und||||chi||0=546  \\$aChinese1False
REPLACE||und||||chi||0=546  \\$aMandarin1False
REPLACE||und||||chi||0=546  \\$aCantonese1False
REPLACE||und||||cze||0=546  \\$aCzech1False
REPLACE||und||||dan||0=546  \\$aDanish1False
REPLACE||und||||eng||0=546  \\$aEnglish1False
REPLACE||und||||fin||0=546  \\$aFinnish1False
REPLACE||und||||fre||0=546  \\$aFrench1False
REPLACE||und||||ger||0=546  \\$aGerman1False
REPLACE||und||||gre||0=546  \\$aGreek1False
REPLACE||und||||hau||0=546  \\$aHausa1False
REPLACE||und||||heb||0=546  \\$aHebrew1False
REPLACE||und||||hin||0=546  \\$aHindi1False
REPLACE||und||||hun||0=546  \\$aHungarian1False
REPLACE||und||||ice||0=546  \\$aIcelandic1False
REPLACE||und||||alg||0=546  \\$aInnu1False
REPLACE||und||||sla||0=546  \\$aInterslavic1False
REPLACE||und||||gle||0=546  \\$aIrish1False
REPLACE||und||||ita||0=546  \\$aItalian1False
REPLACE||und||||jpn||0=546  \\$aJapanese1False
```

Figure 5. Sample of the tasks to populate 008 language, conditional

The task list does not have to do anything clever with the 008 positions because the record data is consistent, so the tasks can search for any instance of ||und|| and replace it with the correct language code depending on the text of the 546. This would also be possible to do in Excel before importing the data, but this method limits the number of manual manipulations we have to do with each new spreadsheet.

Another task copies this new 008 language code into a 041 language field (Figure 6). The final language task in the screenshot on the bottom right amends the 546 text to add 'In [language] with English subtitles', since all of the streaming video collection is accessible in English. If the language of the film is English, making the field data 'In English with English subtitles,' the field is deleted.

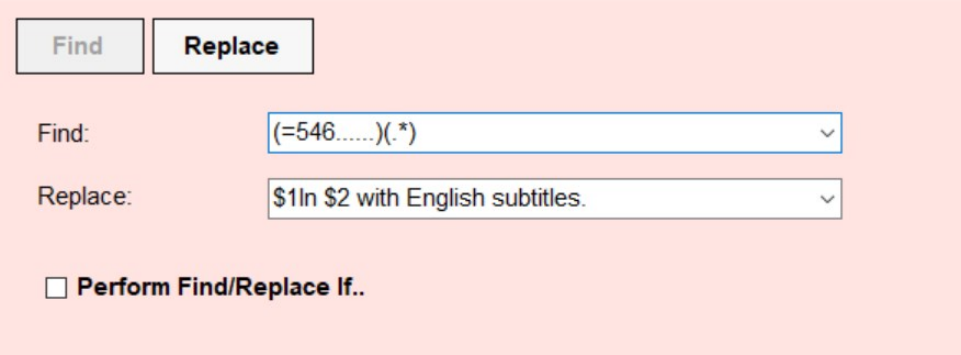


Figure 6. Task to populate language field with subtitle information.

Here's another look at how the task list manoeuvres data out of and into the 008 field (Figure 7). The first task populates the 008 position 18 from the 300 field, but the 008 will only recognise that value if it's a three-digit runtime. So there are two more tasks to add initial 0s if the runtime is only one or two digits.

```
SUBFIELD_EDIT300a008|18||103|1
EDITFIELD008(\d{2})(\|)20$118:3
EDITFIELD008(\d{1})(\|)(\|)200$118:3
```

Figure 7. Three tasks to add the run time from the 300 to the 008, then add leading 0s to

A final task in Figure 8 reverses the order of the names in all 700s to Last, First, relator term. This is not an exact science, as many Chinese and Korean names are already in the Last First order in the spreadsheet and the 511, but it takes less time to make those incorrect at this stage and then correct them manually when we validate the headings. There is potential here to add another task to only reverse the order of the names if the language codes KOR or CHI are not present in the 008, but that would need extensive testing.

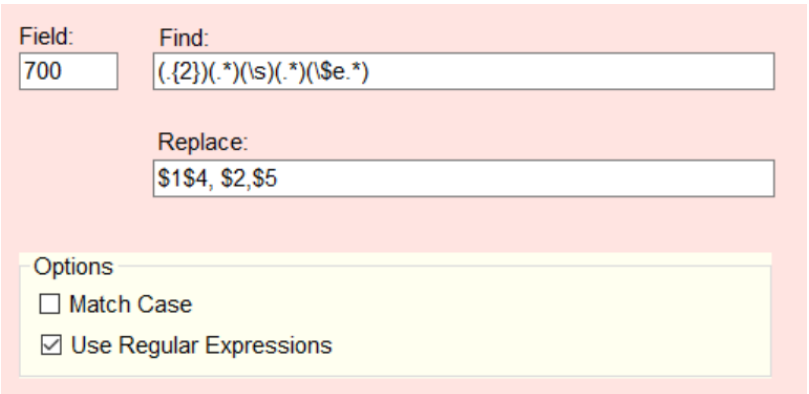


Figure 8. Task to reverse the name order in all 700 fields after the first space.

After running the second task list (Figure 9), you can see our record has effectively doubled as compared to Figure 3. Those accessibility fields are now analysable and also discoverable because they're present correctly in the 341, 347, 532, and also the 655. The relator terms have been added and the 008 has been updated with runtime, year of distribution, year of release, country of distribution, language, form of item, and type of visual material. The 700s have been reversed to Last name First name. And we've also added our local information into the 040 for when this record is shared with other institutions.

```
=LDR  00000ngm a2200000Ia 4500
=006  m\\\\\o\\c\|\\\\\\
=007  vz#uzazuu
=007  cr#cnunnn|uu|u
=008  230831p20231953enk136\|\\\\\|o\\\v|jpn||
=024  7\$a9afb773d-ce2e-59dd-b516-b13c84dada79$2bfi
=040  \\$aUkLeU$beng$erda$cUkLeU
=041  \\$ajpn$jeng
=046  1\$k1953$2edtf
=130  0\$aTōkyō monogatari (Motion picture)
=245  10$aTokyo Story.
=257  \\$aJapan$2naf
=264  \2$a[England] :$bBFI,$c[2023]
=300  \\$a1 online resource (136 mins.)
=336  \\$atwo-dimensional moving image$btdi$2rdacontent
=337  \\$acomputer$bc$2rdamedia
=338  \\$aonline resource$bcr$2rdacarrier
=347  \\$avideo file$2rdaft
=500  \\$aCountry of origin: Japan.
=500  \\$aOriginally released as a motion picture in 1953.
=506  \\$aBFI allows unrestricted access within the UK.
=506  1\$aAccess ends on 02/09/2025.
=511  1\$aChishu Ryu, Chieko Higashiyama, Setsuko Hara.
=520  \\$aA constant fixture in critics' polls, Yasujirō Ozu's most enduring masterpiece, Tokyo Story, is a beautifully nuanced exploration of filial
duty, expectation and regret. From the simple tale of an elderly husband and wife's visit to Tokyo to see their grown-up children, Ozu draws a
compelling contrast between the measured dignity of age and the hurried insensitivity of a younger generation.
=521  8\$aBBFC certificate: U
=546  \\$aIn Japanese with English subtitles.
=588  \\$aDescription based on vendor-supplied metadata.
=650  \4$aAgeing$vDrama.
=650  \4$aFamily$vDrama.
=653  \6$aDrama
=655  \4$aStreaming video.
=655  \7$aFeature films.$2lcgft
=655  \7$aFiction films.$2lcgft
=700  1\$aHigashiyama, Chieko,$eactor.
=700  1\$aRyu, Chishu,$eactor.
=700  1\$aHara, Setsuko,$eactor.
=700  1\$aOzu, Yasujirō,$edirector.
=856  40$uhttp://player.bfi.org.uk/subscription/film/watch-tokyo-story-1953-online
```

Figure 9. Example record after both task lists have been run.

After this task list process is complete—which takes about ten seconds—we do some spot checking to make sure it all appears as it should, and then use MarcEdit's inbuilt tools to validate the 700 and 130 headings, and correct most of them manually (since the BFI-provided names and original film titles, like the country names, are not in Library of Congress authority format). This manual validation process still takes the longest. We leave the subject headings as we're given them from BFI, but indicate that they're local headings (650 \4, as in Figure 9 above). We then create a much shorter and less detailed Delete file which matches on the BFI unique identifier in the 024 to delete films whose access has expired, and finally we send those two files off to BFI to send to their subscribers.

Development and looking forward

All of that development was a huge amount of work to finish before we started providing the records to other institutions at the end of May 2023. But there have also been a few improvements since then as we continue to develop our relationship with BFI.

First, our major pain point was identifying films whose access had been extended. They wouldn't be obvious in the spreadsheet since we add new films based on the access start date, but sometimes older films who had already expired would be extended without the start date being edited. In response to this need, BFI have added an additional column to indicate which access dates have been edited in the past month, and this has made the process of identifying those extensions much more efficient.

Another enhancement we've made is to start adding Library of Congress URIs to our validated 700 entries, thanks to MarcEdit again - there's an option to do this when validating the headings.

At the moment, we receive this spreadsheet of titles in the last week of the month, and have a few working days to turn around the files before they are then emailed out to other institutions. Many suppliers host files like these for institutions to download rather than emailing them out, so there is potential there for streamlining the supplying process. We're also looking for more clarity on how many other subscribing institutions use Ex Libris Alma for their LMS like we do, because we have an opportunity to use Alma's Community Zone to share records, but that is only helpful with a critical mass of Alma users.

And finally BFI offer a package of freely available material in addition to the subscription films, and some subscribing institutions have expressed interest in getting records for those films as well. We're currently in talks with BFI to understand the turnover and demand for those Free titles and may be able to offer this in future.

Impact

The biggest impact on our team locally is relief on our staffing resource. A process that used to take at least a day, and potentially up to four days with a large import, is now maximum a few hours, including the time it takes to validate the Library of Congress authority headings. It's also been an incredibly useful exercise for us, both in terms of a huge stretch project for our team's understanding of MarcEdit and regular expressions, and also in developing our knowledge of streaming video cataloguing standards. We've been able to offer value to our fellow UK HE institutions - at no price increase - who no longer have to create their own records in the painstaking way that we were. Plus, of course, the discount for our own subscription is a bonus.

Most importantly it's had a positive impact for students and other catalogue users, now that the records are provided faster and to a higher standard - especially for things like the accessibility fields, which are much more user-friendly now.

From the BFI side it's also been a useful partnership. We've gotten some really encouraging feedback from them, including from Simone Pyne who is BFI's Senior Business Development Manager: "Our partnership with the University of Leeds has helped us to deliver a much-requested resource by our BFI player subscribing institutions. I have come to learn how crucial MARC records are in aiding discoverability, which is of the utmost importance to us, as our aim is for students and staff to use their BFI player subscriptions to engage with the cultural value of film and support their studies. We didn't have the expertise to create these records in-house, and the insight of the UoL team has been beyond valuable."

We're really pleased to be ambassadors for library metadata standards and for MARC records, and of course also thrilled to have the feedback that the relationship is mutually beneficial.

This project has been useful for us in the Metadata team, but it's also expanded agency within our supplier-vendor relationships for the wider Collections team, and has provided value to our vendor BFI. We're looking forward to future potential applications of similar task lists and future development of these records, whilst celebrating the work that has got us to this point.

The NBK and the UK Distributed Print Book Collection (UK DPC)

Rozz Evans, Head of Collection Strategy, UCL Library Services
Bethan Ruddock, Senior Product Owner, Jisc Library Hub

For many years, the UK's academic library community (including special and national libraries) has been talking about the increasing urgency of a shared approach to ensure the preservation of, and preservation of access to, increasing numbers of scarce print books held in libraries across the UK.

The ongoing move towards digital first provision, pressure on physical spaces, and the challenges posed by ever-increasing demands on staff time and financial resources make this time critical. There is precedent in the form of the UK Research Reserve (UKRR),[1] which addressed the need for a collective approach to the management of print journals and began life as a project in 2007 before transitioning to a 'business as usual' service at the BL in 2019. It has long been agreed that something similar for monographs was needed.

The UK Distributed Print Book Collection (UK DPBC) seeks to address this need and is supported by Research Libraries UK (RLUK) under their 2022 – 2025 strategy, The Library Transforming.[2]

The RLUK's Collections Strategy Network (CSN)[3] Print Storage Group[4] is leading this work, but it is built on the back of many years of discussion, and some significant pieces of work previously completed are listed at the end of this article for further reading. Credit is due to all those involved across the sector that have brought us to this point.

While this vision was endorsed In May 2022 by the RLUK Board of Directors, it must be emphasised that this shared print collection will need to extend far beyond RLUK libraries, with SCONUL, Legal Deposit Libraries (LDLs), special libraries and Jisc all playing a vital role.

What is it?

In a nutshell, it is a nationally distributed book collection based on a minimum number of retained copies, enabling individual libraries to make local decisions about the retention or disposal of their print items which aren't part of the national collection, knowing they can do so without affecting wider long-term access to that content.

The objectives of the UK DPBC are:

- To maximise the value of the shared national collection
- To preserve access to print books at a national level
- To benefit from the shared collection
- To provide awareness of rarity and scarcity within Library collections
- To give confidence to decision making in the local management of print Collections
- To facilitate decisions on space allocation
- To secure 'at risk' print book material
- To provide reliable access to scarce material

---

[1] https://spiral.imperial.ac.uk/handle/10044/1/73162
[2] https://www.rluk.ac.uk/the-library-transforming-strategy/
[3] https://www.rluk.ac.uk/rluk-csn/
[4] The Print Storage group consists of Jane Saunders (University of Leeds), Sandra Bracegirdle (University of Manchester), Hannah Mateer (University of Edinburgh), Joseph Marshall (National Library of Scotland), Michael Williams (University of Cambridge), Sarah Thompson, (University of York), Stuart Dempster (University of Southampton), Rozz Evans (University College London)

Complexities

The aim is simple, and the need is clear, but it there are many complexities, and this is why consulting with stakeholders and making sure concerns are discussed is vital. The following are some of the areas that have been identified – this list is not exhaustive!

- **Accuracy of data**

Clearly accurate, quality, shareable metadata is crucial, and the whole project is dependent on this. Work has been ongoing with colleagues at the Jisc National Bibliographic Knowledgebase (NBK) – of which more later in this article.

- **Number of copies:**

The final agreed number needs to be robust enough to accommodate e.g., lost or damaged copies without being over-cautious and thus cancelling out the benefits of such a collective approach. At time of writing, a number between 5 and 7 copies is looking to be most supported option. We hope to make a final decision soon, following some final data analysis of RLUK libraries as well as some volunteer SCONUL libraries.

- **Preserving access to content**

Legal Deposit Libraries (LDLs) are likely to play a crucial part in terms of retention of copies, given they are already committed to permanent retention for a large percentage of their stock. Questions have arisen in terms of how far loan status needs to be a consideration.

Taking loan status into account adds a layer of complexity which may not be feasible to support. This includes the fact that for most libraries loan statuses are subject to change, the NBK doesn't store loan status data, and adding additional data added to retention statements considerably increases the burden on libraries. On balance it seems pragmatic to ensure the national scarcity threshold is set to take this into account.

We don't expect the UK DPBC to drive ILL usage significantly. It's important to remember that we are talking about safeguarding low use items, so by their nature, it's unlikely that a library would be inundated with requests to borrow. However, access **is** a key issue, and we hope that where possible a holding library will be able to lend (although we recognise it may not always be feasible).

Future digitization initiatives, ongoing developments in mechanisms for sharing resources etc. will all have a part to play in the future.

- **Frequency of NBK updates and use of Retention Statements**

Keeping the data in the NBK as current as possible is vital and support will be given to help libraries to implement this where necessary.

There are guidelines for using retention statements that have been agreed with the community.[5] However, please be aware that these will be revisited over the next 6-12 months to ensure they are meeting the needs of UK libraries, and the UK DBPC work.

Work is ongoing with the NBK to establish methodologies to make it as easy as possible for libraries participating in the UK DPBC to upload retention statements, and to use the Library Hub Compare tool to understand how many copies are available nationally as a routine part of stock editing activities. Ideally retention statements will be added where a decision to retain has been made, but libraries are likely to work at different rates on this, so the key ask will remain to ask libraries not to discard their copy if it is one of *x* or fewer across the NBK.

[5] https://libraryservices.jiscinvolve.org/wp/2020/06/recording-retention-commitments-lhcab-confirms-recommendation/.

- **Physical condition**

There is scope within the retention statement for libraries to add a condition statement if they wish to. However, we don't propose making this mandatory, as many libraries won't have the capacity to do this. We propose that we set the minimum number of copies at a level robust enough to mitigate the fact that some extant copies may be in poor condition (or even missing).

- **Scope**

We don't propose limiting the scope of the printed books included. We feel that this would add too much complexity. We note that for foreign language material in particular this may leave libraries with items, or collections, that they no longer wish to hold. We propose that to take account of this we develop a mechanism for flagging where the holding library cannot retain items it should, under the terms of the UK DPBC, retain. Further consultation on this will be undertaken (but we don't feel this needs to delay the current work).

- **Sustainability**

This is clearly a key benefit for the UK DPBC. There is the potential for all participating libraries to not only save on storage costs and reduce carbon footprints without disadvantaging users.

What will participating in the UKDPBC involve?

It is understood that different libraries will have different capacities to commit and that will be reflected in various levels of membership under an agreed Memorandum of Understanding (MoU). We have access to the UKRR MoU and have considered some interesting examples of similar international projects such as the EAST distributed shared print program in the US.[6]

However, the following draft guidelines give a flavour of what will likely be expected.

1. To **agree to retain** a copy of a print book if there are fewer than *x* copies held within the NBK.
2. To **add retention statements** to the metadata for items you have committed to keep for the foreseeable future or until a set date.[7]
3. To **upload metadata on a regular basis** including retention statements to the NBK.
4. To **facilitate access** to items with a retention statement, ideally by ILL where this is feasible.
5. To identify **last copies** and retain where possible (work under way to clarify what to do if a library is physically unable to commit to retain).

Governance and administration of the UK DPBC is being considered and further information will be available in due course.

[6] https://eastlibraries.org/about-us/about-shared-print/

[7] https://libraryservices.jiscinvolve.org/wp/files/2020/06/LHCAB-Retention-Statement-Final-0620.pdf

Role of NBK and metadata

We believe that Jisc and NBK will play a key role in the UK DPBC. Success will rely on good quality, reliable holdings date: consistent use of retention statements; shared commitment; minimal administrative burden for individual libraries; continued development to support.

How Library Hub is supporting the UKDPBC

Jisc currently runs three Library Hub service: Library Hub Discover, a freely available service for resource discovery (discover.libraryhub.jisc.ac.uk); Library Hub Compare, which facilitates collection development and management (compare.libraryhub.jisc.ac.uk); and Library Hub Cataloguing, a MARC records download service (cataloguing.libraryhub.jisc.ac.uk).

The Jisc Library Hub services are built on a key piece of infrastructure, the National Bibliographic Knowledgebase (NBK). The NBK is an aggregation of 51,612,311 records created from 152,616,854 records contributed by 204 institutions. These are a mixture of UK HE and specialist research institutions, plus the national libraries and the library of Trinity College Dublin. A full list of contributing libraries can be found at https://discover.libraryhub.jisc.ac.uk/about/libraries/.

Bibliographic records from these libraries are supplied to the Library Hub team, usually in MARC format, and they are then loaded into the NBK data lake. Out of this data lake, a deduplication process is run which merges records for the same item. This dataset is used to provide the Discover and Compare services, while an unduplicated database providing permissibly-licensed MARC records as provided by the libraries is used for the Cataloguing service.

This database is kept updated by regular data uploads from contributing libraries; however the lag inherent in the system mean that the NBK is never an entirely comprehensive or current reflection of the UK national collection.

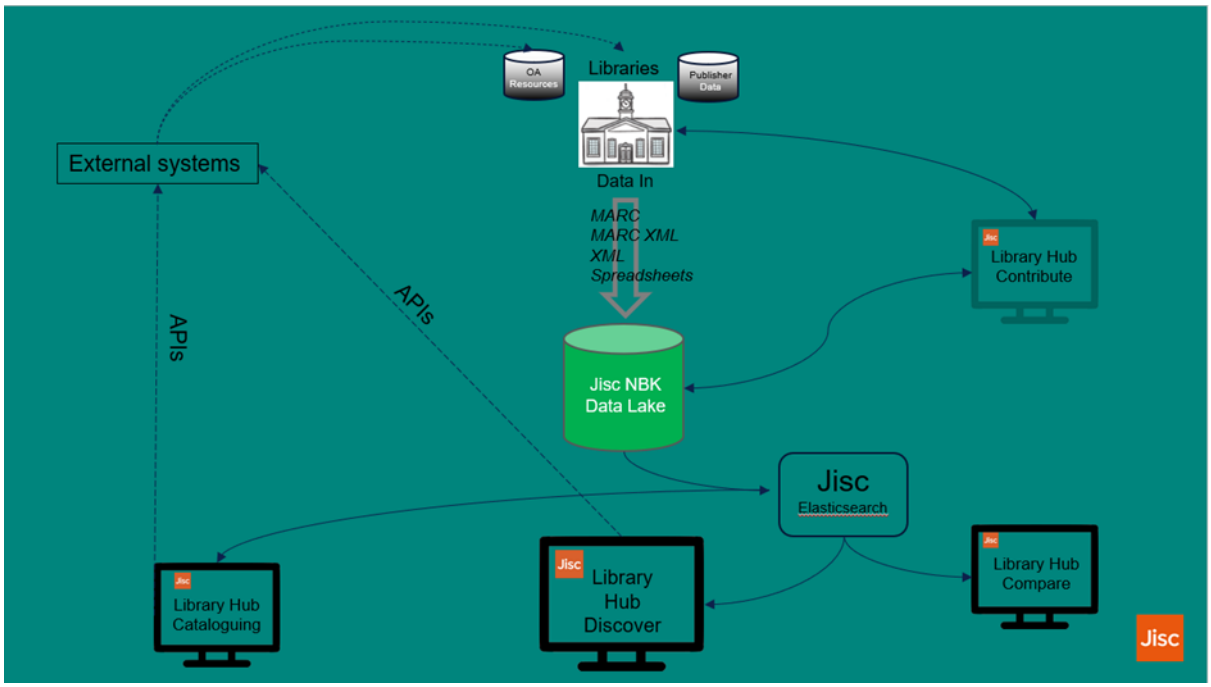


Figure: an overview of the architecture of the Library Hub services

As the NBK is a unique aggregation of the holdings of UK libraries, we are very well-placed to be a data source for the rarity assessment work being undertaken as part of the UKDPBC work.

Following initial conversation with the RLUK CSN members in 2022, we produced a spreadsheet giving an overview of the number of print books held by 10 or fewer libraries (9 or fewer, …, down to 1 holding library) in the UK, with breakdowns by consortia and individual library. An example section of the spreadsheet can be seen below.

	Totals		Consortia: at least 1 of the holding institutions is in the consortium				
No. holdings institutions	Result	At least 1 record includes NBK-R	RLUK	SCONUL	SCONUL NOT RLUK	Legal Deposit Libraries	Other (not SCONUL or RLUK)
Total deduplicated records in the NBK	50341867	82314	40445825	44481163	4801890	27233665	5107978
Total deduplicated print book records in the NBK.	33217075	80964	28402962	29939782	2092010	20814168	2722216
No. of records with 10 or fewer holdings institutions	31524309	52916	26712578	28249373	2091981	19162186	2721836
No. of records with 9 or fewer holdings institutions	31344460	51185	26532735	28069415	2091866	18994016	2721836
No. of records with 8 or fewer holdings institutions	31132169	49217	26320640	27857074	2091620	18792396	2721836
No. of records with 7 or fewer holdings institutions	30852371	47069	26041719	27577236	2090703	18529767	2721836
No. of records with 6 or fewer holdings institutions	30453041	44453	25643107	27177962	2090041	18154166	2721836
No. of records with 5 or fewer holdings institutions	29745526	41819	24936427	26470154	2088910	17476084	2721834
No. of records with 4 or fewer holdings institutions	28955544	38966	24149415	25680085	2085828	16742421	2721827
No. of records with 3 or fewer holdings institutions	28023153	35691	23230237	24747957	2072716	15937647	2721660
No. of records with 2 or fewer holdings institutions	26822406	31374	21853117	23351125	2051292	14784826	2719644
No. of records with 1 holding institutions	23608892	24084	18960607	20392338	1965634	12586242	2684462

Investigation of these figures soon showed that they were not an accurate representation of the print holdings in UK libraries, with many libraries reporting that the figures provided grossly overstated their print book holdings. The issue was that records in the Library Hub are deduplicated regardless of format. Library Hub Discover was designed with the aim of satisfying the end-user need of 'I want this book, how can I get it?'. This means that records for print and e versions are merged into the same record. When you filter the deduplicated dataset by 'print', you won't get any results where all of the items are electronic, but you will get mixed print and e as well as all print. For instance, this record...

10 Holding libraries

British Library

University of East Anglia Library

Institute of Advanced Legal Studies Library

LSE Library

National Library of Scotland

National Library of Wales / Llyfrgell Genedlaethol Cymru

University of Nottingham Libraries

University of Oxford Libraries

UCL Library Services

University of Warwick Library

...was being counted as ‘print, with 10 holding libraries’, despite 1 of them being electronic.

The Library Hub Compare service already has a deduplication system which avoids this mixed-format problem, but because of system demands that can handle a maximum of 50,000 records at a time; not feasible for work across the entire dataset.

What we have done is to create a new development version of the NBK database with a new index which only counts the number of print holdings, and excludes other formats from the count. This also enable us to provide more nuanced reports to the individual libraries, only including consolidations where they hold a print copy.

We are currently in the process of creating a set of dashboard visualisations which we will supply to the RLUK CSN members, along with the individual libraries who are taking part in the initiative.

Next steps after this will be to create a report allowing libraries to ingest and examine a list of their holdings which meet the rarity threshold. Initially these lists will be generated by the Library Hub team, with the intention of eventually making them self-service through the Library Hub Compare service.

While there will always be some errors based on either scarce or inaccurate metadata, especially where the format has not been correctly identified in the record; or rare misconsolidations (from over 45 million searches a year on the Discover we have on average fewer than 10 misconsolidations reported to us), we are confident that the data will allow for confidence in an accurate assessment of print rarity in libraries across the UK, and are pleased to be supporting this important work.

References and Further Reading:

Theo Stubbs et al., *UKRR Final Report*, August 2019, DOI: https://doi.org/10.25561/73162 (accessed 13/11/23)

The Library Transforming: RLUK’s Strategy for 2022 – 25 https://www.rluk.ac.uk/the-library-transforming-strategy/ (accessed 22/11/23)

RLUK Collections Strategy Network https://www.rluk.ac.uk/rluk-csn/ (accessed 22/11/23)

White Rose Library staff, Understanding Collections Overlap: An Investigation Into White Rose Libraries Collections Using Collection Management Tools, Final Report, July 2017, https://blog.ccm.copac.jisc.ac.uk/wp-content/uploads/sites/13/2017/09/Understanding-collections-overlap-final-report-July-2017-including-appendices.pdf (accessed 13 November 2023)

Constance Malpas and Brian Lavoie, Strength in Numbers: The Research Libraries UK (RLUK) Collective Collection, 2016, https://www.oclc.org/research/publications/2016/oclcresearch-strength-in-numbers-rluk-2016.html (accessed 13 November 2023)

Information Power, Feasibility Study on Monographs, Final Report, June 2017, https://www.sconul.ac.uk/sites/default/files/documents/Feasibility_Study_on_Monographs_final_report_June_2017_0.pdf (accessed 13 November 2023)

Theo Stubbs, An Extension of UKRR Into Low-Use Monographs: Does Appetite Exist?, November 2018, http://hdl.handle.net/10044/1/69379 (accessed 13 November 2023)

EAST – Eastern Academic Scholars Trust distributed model shared print program https://eastlibraries.org/about-us/about-shared-print/ (accessed 22 November 2023)

Library Hub Community Advisory Board – Retention Statement (Final) https://libraryservices.jiscinvolve.org/wp/files/2020/06/LHCAB-Retention-Statement-Final-0620.pdf (accessed 22 November 2023)

Library Hub Discover https://discover.libraryhub.jisc.ac.uk/ (accessed 22 November 2023)

Library Hub Compare https://compare.libraryhub.jisc.ac.uk/ (accessed 22 November 2023)

Library Hub Cataloguing cataloguing.libraryhub.jisc.ac.uk/ (accessed 22 November 2023)

NBK Contributing Libraries https://discover.libraryhub.jisc.ac.uk/about/libraries/ (accessed 22 November 2023)

Updating the Wessex Classification Scheme for UK health libraries: a case study in improving inclusion and diversity in a specialist classification scheme

Jason Curtis, Librarian, Shrewsbury and Telford Hospital NHS Trust
Lotty Summers, Library Assistant, University Hospitals Birmingham NHS Foundation Trust

Introduction

The Wessex Classification Scheme[1] was created in the early 1980s by healthcare librarians in the former Wessex Regional Health Authority area in the South West of England, with the aim of improving the cataloguing of items relating to UK healthcare practice.

The scheme consists of two parts. The first is the classification schedule itself, which was loosely based on the US National Library of Medicine (NLM) classification.[2] The second part is an annotated subject index that is based on a subset of the NLM Medical Subject Headings (MeSH) scheme[3] but with Anglicised spelling and some local additions.

The major differences between the NLM and Wessex classification schedules are around psychology and nursing. The librarians of the Wessex region created a new schedule for psychology, and classified books on nursing specialties within the specialty. For example, a book on cardiac nursing is shelved with cardiology books.

The Wessex name is something of a misnomer now as the scheme is widely used in healthcare libraries across the UK, both inside and outside the NHS. Some NHS regions use it exclusively across their library services and have adopted the subject headings list as an authority file in regional library management system consortiums such as SWIMS (South West), HeLM (Midlands), and the upcoming North West and Yorkshire and Humber consortia.

Maintaining the scheme

Although the scheme has gone through several revisions, there had been no major updates since 2013, so the Wessex Classification Scheme Oversight Group was formed in September 2022 with the support of Health Education England (HEE) (now part of NHS England). The group brings together knowledge and skills from across UK health library networks to improve the scheme, whilst offering the opportunity for participants to develop professional skills in classification and subject indexing. It also provides an opportunity to network widely, helps ensure the longevity of the scheme, and shares the maintenance work.

One of NHS England's stated goals is to 'mainstream equalities, diversity and inclusion into every team, function and role'[4] and their support for updating the Wessex scheme is helping to meet this aim, by creating a higher quality, more inclusive experience for library users.

Up until the formation of the Oversight Group, the Wessex scheme was very much a product of the South West NHS region and was maintained for many years by a small cataloguing group. Changes to the scheme, particularly the subject headings, were not encouraged where these diverged from NLM MeSH due to the difficulty of updating the schedules. These schedules took the form of a series of discrete document files and any change might affect cross-references and require amendments to multiple files, leaving a large margin for error or missed references.

---

[1] SWIMS Network (2023). Wessex scheme. Available at https://swimsnetworknhs.uk/how-to-guides/cataloguing/wessex-scheme/
[2] National Library of Medicine (2023). NLM classification. Available at https://classification.nlm.nih.gov/
[3] National Library of Medicine (2023). Medical Subject Headings. Available at https://www.nlm.nih.gov/mesh/meshhome.html
[4] Health Education England (2023). Creating a new NHS England. Available at https://www.hee.nhs.uk/about/how-we-work/your-area/london/london-news/creating-new-nhs-england

However, there were small numbers of local headings introduced into the subject headings index, identified within the index as 'Wessex MeSH'. These tended to be terms relating to NHS organisational structures or UK terminology.

Formation of the group

In the summer of 2022, a call was put out for health library staff interested in being part of a nationwide group to update the Wessex scheme. Members were recruited via health library mailing lists such as LIS-Medical, and through communications from HEE. The new Oversight Group collaborates via the FutureNHS platform, and meetings are held virtually via Microsoft Teams.

Opening up the Oversight Group to library staff across England means that there is a greater pool of knowledge and enthusiasm to draw on. Around 30% of the group are para-professional staff, with the rest holding various librarian roles including clinical and outreach librarians, along with a small number of library service managers.

Initially, members were asked which parts of the scheme they felt needed updating the most and sub-groups were then formed for LGBTQ+ issues and gender identity (known as the Pride sub-group), Ethnicity and Race, and Learning Disabilities and Neurodiversity (the LDN sub-group).

Pride sub-group work

Since the last revision of Wessex in 2013, awareness and literature on the topic of LGBTQ+ issues has increased. The sub-group (some of whom identify as LGBTQ+) felt strongly this was an opportunity to educate about LGBTQ+ issues and care requirements.

The sub-group began their work by investigating the HQ classification schedule which covers sexuality, relationships, marriage, and the family. As part of the process, it was found that criminal aspects of sexual behaviour (such as paedophilia and incest) were located within the same section as topics relating to sexual orientation and gender diversity. The sub-group believed this perpetuated a stereotype by linking these topics and suggested they move from HQ to WM (Psychiatry, Mental health) in the classification.

Next, subject headings related to the LGBTQ+ community, gender identity, and gender affirming care were reviewed. This found terms that were no longer used or could be considered offensive, derogatory, or prejudicial. As a result, some terms were removed completely (for example 'hermaphroditism'), and other terms were updated (for example 'gender affirming surgery' rather than 'transsexual surgery'). Resources such as Homosaurus[5] were used to clarify the best terminology and provide scope notes in the absence of an NLM MeSH definition.

Proposed changes were shared with LGBT+ staff networks and their feedback provided a useful perspective. The work of the sub-group is now complete, and changes were published to the scheme in June 2023 in time to celebrate Pride month.

Ethnicity and race sub-group work

The focus of the Ethnicity and Race sub-group is to ensure that Wessex subject headings are inclusive and current as well as removing terms more appropriate to an American context. In addition to terms used to describe ethnic groups, related topics such as immigration, traditional medicine, homelessness, and health inequalities were also considered.

---

[5] Homosaurus (2023). Homosaurus: an international LGBTQ+ linked data vocabulary. Available at https://homosaurus.org/

As the subject matter runs throughout the classification scheme it was not possible to interrogate specific classmarks. Instead, subject headings were transformed into one spreadsheet covering all the A-Z documents, which enabled the sub-group to search for potential terms for review. Appropriate resources were consulted before making recommendations to the Oversight Group.

The sub-group faced challenges, with members of the group leaving and the original chair stepping down. Additionally, the membership of the group does not reflect lived experience of the issues concerned and attempts to contact relevant NHS Trust Networks were unsuccessful.

Finding times to meet proved difficult and so the sub-group mostly worked on shared documents via the FutureNHS platform, eventually collating a document of proposed new subject headings (such as Indigenous Peoples), terms for removal (such as African Witch Doctor) and for rewording (for example 'Medicine, oriental' to 'Medicine, traditional').

Learning Disabilities and Neurodiversity (LDN) sub-group work

The term 'neurodiversity' encompasses a range of differences in brain function and behavioural traits, as well as learning disabilities, and it is estimated that 1 in 7 people in the UK are neurodivergent.[6] The Wessex scheme currently has no subject heading for 'neurodiversity' or for neurodivergent people, yet there is a growing body of literature on neurodiversity available in healthcare libraries.

The LDN sub-group is looking at the language used in the subject index and considering some broader changes to the classification schedules. Some members of the LDN sub-group have lived experience of neurodiversity and are able to bring their experience to this work.

There are several problems with the way Wessex is currently structured regarding neurodiversity, particularly the assumption that neurodiversity is a childhood-only issue. For example, the subject heading for Asperger syndrome (itself an outdated term) specifically describes it as a 'childhood disorder' and suggests that the classmark is within the paediatric section.

Another problem is multiple classmarks for the same topic. An example is dyslexia, where there is the choice of using a classmark within paediatrics, within mental disorders if it is manifested in adulthood, within neurology if it is acquired (for example due to brain injury) or within education if it relates to study skills and teaching methods. This means that books on dyslexia could be shelved in four entirely different places in the library.

The LDN sub-group looks at neurodiversity not through a medical model lens (where for example these conditions might be described as 'mental disorders') and instead seeing it through a social model lens. It has been suggested that changing the structure of the classification scheme and providing up to date subject terms with definitions has an educational function both for library staff doing the classification and cataloguing of materials, but also for library users browsing the shelves or searching the catalogue, and will hopefully improve understanding of neurodiversity.

The sub-group aim to create a new classmark within the Wessex scheme that sits between WLM (psychology) and WM (psychiatry), then to move materials on neurodiversity and learning disabilities to this new sequence, prefixed with WLN. This would bring together neurodiversity books from disparate sections, and remove these books from sections that are concerned with mental disorders.

There are drawbacks to this approach; the work involved in creating an entirely new class would be time consuming and there are implications for specialist mental health libraries that may have large numbers of books needing reclassification.

[6] British Dyslexia Association (2019). New Guidance: Neurodiversity in the Workplace. Available at https://www.bdadyslexia.org.uk/news/new-guidance-neurodiversity-in-the-workplace

The group has faced membership struggles and has lacked professional expertise since the group was almost entirely composed of para-professional staff. However, after a call for new members, the LDN sub-group has some new members, including two librarians as co-chairs.

The future

A year in, the Wessex Oversight Group has much to show for this important work, and there is more to come in the next phase.

Firstly, the Wessex Lead has met with EBSCO to discuss how we might work together to update subject headings in the CINAHL database.[7] EBSCO have recognised the need to update some of their terminology since their subject headings, like Wessex, are often drawn from NLM MeSH terms. We've shared the work of the Pride sub-group and this has led to some changes to the CINAHL thesaurus. We'll be sharing work that we do in future, so that they can see the changes we make, and they will make us aware of their latest updates.

Secondly, there is work to move the Wessex scheme into a thesaurus management tool called PoolParty. This will provide links between the subject index terms and the classification schedules and will make it easier for cataloguers to interrogate the scheme. PoolParty will ease the updating of subject headings, particularly any cross-references and related terms, and this should facilitate the use of alternative terms to those in the NLM MeSH scheme where it is felt we can promote more inclusive language going forward.

The move to a thesaurus management tool will also facilitate a move towards being hosted on the NHS England website encouraging the scheme's wider use by making it more accessible on a national NHS platform.

Lastly, new sub-groups have been set up to look at several areas, including a group to look at suggestions from users of Wessex in the library community. Going forward we foresee that rather than major new editions of the scheme, there will be smaller updates published on a quarterly basis, allowing the scheme to be more flexible in adapting to changing terminology and healthcare practice.

[7] EBSCO (2023). CINAHL Database. Available at https://www.ebsco.com/products/research-databases/cinahl-database

TIMELINE OF WESSEX CLASSIFICATION SCHEME

1951

First edition of National Library of Medicine classification scheme.

1960

Introduction of the Medical Subject Heading (MeSH) scheme.

1987

First edition of the Wessex Classification Scheme and subject heading index.

1999

Third edition of the Wessex Classification Scheme is published.

2004

A survey indicates that the scheme is widely used across the UK.

2013

Fifth edition published. The last major revision, but small updates published after.

2022

National survey takes place, followed by the formation of a new Oversight Group.

2023

First significant revisions published in June, on LGBTQ+ and gender identity topics.

Introducing the Wikidata Thesis Toolkit

Helen Williams, London School of Economics Library
Ruth Elder, University of York Library

The Wikidata Thesis Toolkit builds on the foundation of work initiated by Helen Williams (London School of Economics Library) in 2019 as a Covid "lockdown" project. This was prompted by a growing interest in Wikidata as a topic of conversation within the Metadata community, and led to her work on the development of a process to upload pre-existing theses metadata into Wikidata. Subsequent work by Ruth Elder (University of York Library) to develop a similar process flow at York added additional refinements, and the resulting collaboration between the two libraries resulted in the production of the Wikidata Thesis Toolkit.

Wikidata is described as a structured database operating as the central data store for all Wikimedia projects. It is a free and open knowledge base containing multilingual data that can be read, edited, and re-used by humans and machines, supporting global access to information.[1] The challenge was to take advantage of these beneficial attributes of Wikidata in order to develop a sustainable process to input theses metadata (already held in digital institutional repositories or datasets) into Wikidata. This enables the promotion of institutional original research to the widest possible audience through signposting back to the repository.

With competing demands and limited resources academic libraries need to be able to justify investing staff time and effort in developing Wikidata skills to promote resources in general, and institutional doctoral theses metadata specifically. The investment is justified by its support of library and institutional strategic priorities around enabling an "open as possible" approach to accessing research outputs. In linking the open content of digital repositories into Wikidata:

- scholarly content becomes more widely accessible and visible
- the role of the institution as a provider of open knowledge to local and global audiences is promoted
- the creation of unique identifiers for research outputs enables the institutional content (and the entities within it) to become part of the Linked Open Data ecosystem
- search engine results are populated with a fuller picture of globally available data because Google Knowledge Graphs, digital assistants, and Wikipedia infoboxes are all populated, in part, with information harvested from Wikidata.

In addition, many libraries will be looking to expand the work of their Metadata teams beyond traditional cataloging, and to develop staff skill sets to future-proof roles. Working with Wikidata is an accessible approach to introducing Linked Data work and expanding the range of staff digital skills, confidence, and experience.

Based on the experience of their learning journey, the authors collaborated to create the toolkit, which is created with the aim of reducing the development burden for other institutions looking to establish similar projects. It is designed as a guide, rather than a step by step handbook, and is presented with best intent, reflecting the skills, knowledge, and experience of the authors at the time of creation. The toolkit is available through a Wikidata project page, and has been shared with the wider academic library and Wikipedia communities. It supports the open research agenda by providing a method and workflow to help institutions promote their doctoral dissertations to the widest possible audience by increasing their visibility and accessibility. As the amount of doctoral research metadata in Wikidata grows the potential of surfacing unexpected connections and relationships increases, meaning the data can be explored in new ways, beyond institutional silos, to make sense of combined cultural heritage. In practical terms the toolkit includes the following:

[1] Wikidata, https://www.wikidata.org/wiki/Wikidata:Main_Page, accessed October 10 2023

- resources to introduce the basics of Wikidata and provide foundational knowledge (this takes 30-60 minutes and aims to be a low barrier starting point)
- links to further reading
- guidance on setting up an account
- practical/manual tasks to develop basic familiarity in editing Wikidata
- process overview
- steps to add metadata to Wikidata
    - data preparation
    - editing in OpenRefine
    - reconciling institutional names with Wikidata
    - creating Qids for individuals and thesis titles
- instructions to link theses to external identifiers
- guidance on creating links to theses from Wikipedia pages
- methods for using SPARQL to visualise data in ways not usually possible via an institutional repository.

The toolkit also provides guidance on measuring the impact of uploading thesis metadata to Wikidata. In the university environment the word 'impact' is often understood in its Research Excellence Framework context as 'an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia'.[2] In order to avoid confusion when promoting the Wikidata thesis project in a Higher Education setting it has been more appropriate within LSE's internal environment to refer instead to 'reach and engagement'.

The involvement of library colleagues in a Wikidata project provides immediate value through the development of digital skills, the opportunity to contribute to the broader network of Linked Open Data, and the promotion of unique and distinctive collections beyond the usual library silos. These benefits can be highlighted to institutional research and PhD teams, and alumni, to raise the profile of the library and demonstrate the value of metadata in ensuring that the library's unique content can be understood in the semantic web environment.

Senior leadership teams, however, will also want to see concrete evidence demonstrating the benefits of investing time and resources in this work. This evidence is not necessarily readily available. The objective of adding library metadata into Wikidata is to enable other information sources (such as search engines) to access Wikidata as a source of authoritative, referenced, structured data, indirectly enriching the wider data ecosystem. This means that measuring the direct outcomes of a Wikidata project, rather than recording basic statistics on the volume of data created or updated, can be challenging.

Many institutions are keen to benchmark with comparator organisations, so institutional rankings for doctoral theses in Wikidata can be useful, though it is important to note that these will change over time, and that for smaller institutions who simply do not have large numbers of PhD students, these rankings are not taking any account of theses produced per size of institution so they might be less valuable. This result from the Wikidata SPARQL query service provides a count of doctoral theses by institution,[3] while this dashboard includes a count of all theses types by institution and compares the completeness of the metadata available in Wikidata.[4]

LSE has investigated various measurements to review the reach of the project, including:

---

[2] UK Research and Innovation, *How Research England supports research excellence,* https://www.ukri.org/who-we-are/research-england/research-excellence/ref-impact/#contents-list accessed October 10 2023
[3] https://w.wiki/jwZ
[4] https://www.wikidata.org/wiki/Wikidata:WikiCite/Theses_by_institution

- Downloads from the institutional thesis repository, LSE Theses Online.

Increase in downloads

Total downloads 2021 **16% higher** than 2020

Total downloads 2022 **14.2% higher** than 2020

- Figures from Google Analytics

Google Analytics

Overall **250%** increase in users referred to LSETO from Wikipedia between 2019/20 and 2021/22

Contextually, in terms of total referrals into LSETO from Wikipedia:

2019 (pre project) 1%

2020 (project begins) 3%

2021 (project complete) **13% - still consistent Sept 2023**

- Twitter mentions

Twitter - etheses.lse.ac.uk

Nearly doubled in the time period reviewed

Paid resources could be used to analyse more extensively

38 mentions Feb - May 2020

74 for same time period 2021

- Extension of institutional names in Wikimedia

Increase in author/supervisor data in Wikimedia

2019: just 23% of LSE authors and supervisors existed in Wikidata

2023: 100% of LSE authors and supervisors are represented in Wikidata

Just 7% of authors and supervisors have a Wikipedia page - highlighting the quantity of unique data added which can now be used by search engines and Wikimedia editors

Wikidata is based on a structured format to describe entities, in that there are limited, prescribed ways of describing any item. For example, there are a finite number of ways of describing an author, or a dissertation, and it is not possible to add a freehand or self created description. In this way Wikidata is described as structured data.

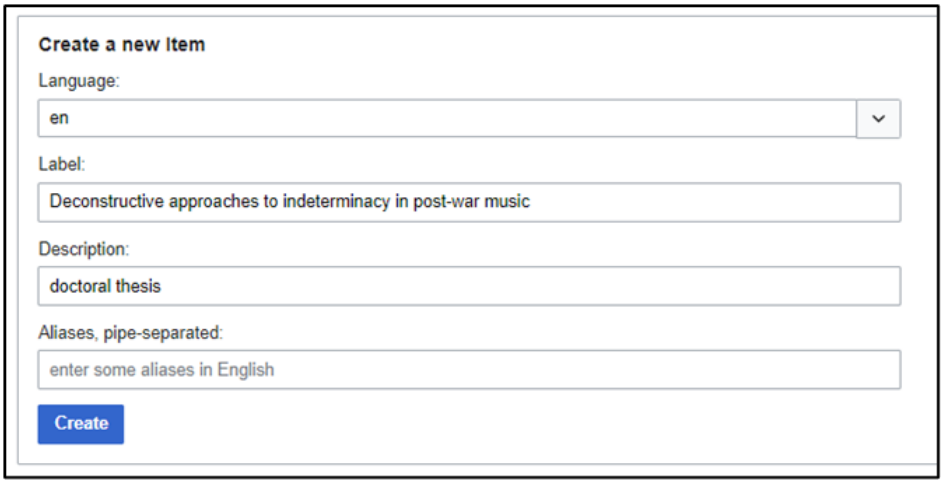
In addition to recording items Wikidata also expresses the relationships between those items. For example, a thesis entry can be linked to an author entry, the author entry can be associated with an awarding institution, and to a doctoral advisor. Each of these entities will have their own individual item entries and hence a Qid in Wikidata.

An accessible initial approach to this work for those who are unfamiliar with using Wikidata is by creating individual manual entries for theses and their authors. This helps to become orientated in Wikidata, make and correct mistakes, build confidence, and understand how the different elements work together, prior to moving onto bulk edits and uploads.

An outline of this process is set out below.

- Set up a Wikidata account and log in.
- Check to see if there is an existing entry for the specific thesis title by entering in search box on Wikidata front page.

- Select Create a new item (at the left of the screen).
- Enter thesis title and description as shown below.



- By selecting Create, this produces a minimal record of the item and a brief description.

At its most basic the structure of Wikidata is based on linking three pieces of information: an item, a property of the item, and a value associated with the property. In this example the item is the thesis titled "Deconstructive approaches to indeterminacy in post-war music." Each item or entity is given a Q number or Qid which is its unique identifier within Wikidata.

The item/property/value combination is referred to as a triple, and a triple constitutes a statement. Multiple statements are stored within a single item. The greater the number of statements within an item record, the more comprehensive, informative, accessible, and searchable the record (through linking to other relevant Wikidata entries).

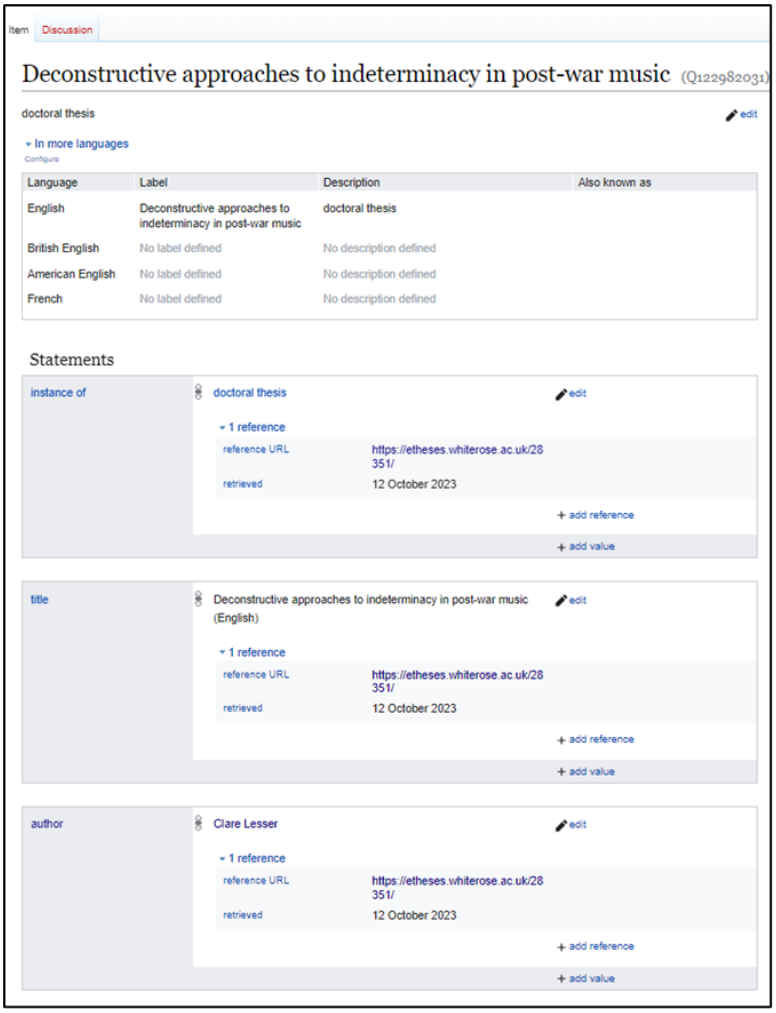
Additional statements combine to form a schema (as shown in the table below) and should be added to the basic record to make it more informative and to connect with other entities in Wikidata.

Item (Thesis title)	Deconstructive approaches to indeterminacy in post-war music	
Description	Doctoral thesis	
Statements **(Item + Property + Value)**		
Item	**Property**	**Value**
Deconstructive approaches to indeterminacy in post-war music	Instance of P31	Doctoral thesis Q187685
Deconstructive approaches to indeterminacy in post-war music	Title P1476	Deconstructive approaches to indeterminacy in post-war music
Deconstructive approaches to indeterminacy in post-war music	Author P50	Clare Lesser Q122982224

Deconstructive approaches to indeterminacy in post-war music	Dissertation submitted to P4101	University of York Q967165
Deconstructive approaches to indeterminacy in post-war music	Language of work or name P407	English Q1860
Deconstructive approaches to indeterminacy in post-war music	Publication date P577	2020
Deconstructive approaches to indeterminacy in post-war music	Full work available at URL P953	https://etheses.whiterose.ac.uk/28351
Deconstructive approaches to indeterminacy in post-war music	On focus list of Wikimedia project P5008	UniversityofYorkThesisProject Q1145883936

The addition of a reference to a statement is good practice, citing where information was sourced from and authenticating it. In this thesis example this would be in the form of referencing the thesis URL (linking to thesis entry in the digital repository) and confirming a retrieval date.

A search on Wikidata for the thesis title will now return the entry created above with a unique identifier of Q122982031. Further statements or identifiers can be added to the record at a later date and/or by any other Wikidata editor.



This structured method of storing data means that Wikidata can be queried using its built-in SPARQL query tool. SPARQL can appear complicated to beginners, and the authors recommend borrowing and editing queries written by other people to assist with learning. Multiple queries can be found on LSE's Wikidata Thesis Project page to support this.[5] By clicking on the links to any of the data visualisations on this page users can select to edit the SPARQL and substitute LSE's Qid or project Qid for that of another institution. These queries link institutional metadata on Wikidata with all the other data already existing in Wikidata and consequently make connections in ways that have not previously been easily discoverable or visible, such as:

- awards won by thesis authors and supervisors
- institutions holding their archives
- their employers (in list format, or visualised by location on a map)
- educational establishments attended
- authors and supervisors with a Wikipedia page
- relationships and chains between authors and supervisors
- image grids and Histropedia timelines
- thesis subject data
- thesis citation data.

It is important to note that an individual institution can be confident of the completeness of their institutional data, once it has been contributed to Wikidata, but the existing data these queries link with in Wikidata is not necessarily complete. This does not diminish the usefulness of the data for search engines, but it does need to be made clear to anyone making use of the data for research purposes.

For both authors Wikidata has proved an enjoyable, delightful, and satisfying challenge allowing the development of new skills, the contribution of unique institutional data to Wikidata, and the increased visibility and accessibility of this content. The Wikidata Thesis Toolkit is presented with the intention of helping other librarians to join this journey, alongside other sources of Wikidata guidance and support. These include:

- Wikidata discussion pages
- WikiEdu (https://wikiedu.org/)
- OpenRefine Community (https://openrefine.org/community)
- LD4 Wikidata Affinity Group (https://www.wikidata.org/wiki/Wikidata:WikiProject_LD4_Wikidata_Affinity_Group)

Support and guidance at the global level has been instrumental in helping both authors progress their projects, yet there is also an awareness that being able to discuss questions and challenges within a local community in the same time zone can be of value. To this end it is hoped that the Wikidata Thesis Toolkit will be a starting point to a growing community of practice among UK Higher Education and GLAM institutions (Galleries, Libraries, Archives and Museums) who are interested in developing Wikidata projects and sharing experiences with one another. To support this a new WikiProject, entitled WikiProject_UK GLAM Wikidata Projects, was launched at the conference.[6] This aims to be a space where those working with Wikidata in the UK GLAM sector can share workflows, reduce development burdens through shared communication, and spot opportunities for collaborative projects and exploring links between sets of data. This space will be a community-led initiative and by using Wikidata as its forum anyone can edit the project page, link to projects, and initiate discussion. Please do engage with the project page, link to useful resources, share projects or pre-project planning, and use the talk page to ask questions and discuss relevant topics.[7] It is hoped this space will grow, and provide details of other innovative and exciting Wikidata work across the UK, with theses and beyond.

---

[5] https://www.wikidata.org/wiki/Wikidata:WikiProject_LSEThesisProject
[6] https://www.wikidata.org/wiki/Wikidata:WikiProject_UKGLAMWikidataProjects
[7] https://www.wikidata.org/wiki/Wikidata_talk:WikiProject_UKGLAMWikidataProjects

UK NACO funnel : progress, obstacles, and solutions - a quick update

Martin Kelleher, University of Liverpool

Abstract

This Lightning Talk will provide a quick update on latest progress with the now established UK NACO (Name Authority Control Programme) Funnel, which allows participating institutions to contribute to Library of Congress / PCC (Program for Cooperative Cataloguing) authority control. The presentation will include a brief summary of the purpose of the funnel, details of latest expansion, problems and solutions with data submission software, and further plans and collaborations.

Introduction

This is just a quick update on the progress, problems, and solutions to problems we've had with the still nascent UK NACO Funnel, after reporting it established in the CILIP MDG Autumn Seminars in 2022.

Progress – building the funnel

For those who don't know what a NACO funnel is, a NACO Funnel is an administrative arrangement by which multiple institutions can contribute to and edit name authority records in the Library of Congress (LoC) Name Authority File (NAF). Contribution can be managed on an individual institutional basis, but this requires a commitment to a level of authority record production many UK institutions would struggle to justify or accommodate, so access through a funnel is a highly attractive alternative, allowing contribution without problematic quota requirements.

The NACO Funnel was established in 2022, based around 4 institutions, which were the University of Liverpool, Sheffield Hallam University, University of Leeds and Aberystwyth University. It is the first UK NACO funnel to have been successfully established. Since the initial establishment of the funnel, it has continued to grow, and now there are 7 institutions (ish!) who have NACO trained cataloguers, the original 4 institutions being joined by Cardiff University, University of Leicester and Solent University. I'll explain the "ish" later….

We've got a JISC email list to manage communications within the funnel, which has grown to having 22 subscribers, between member institutions and some other institutions looking to join the funnel.

OCLC – to have or have not

It's not been plain sailing, however, and there have been a few issues we've had relating to the funnel. Already something of a concern while the funnel was being developed, the issue of OCLC membership had been something that had affected many since the falling-through of the JISC deal that had been in the process of being established, in July 2022.

There had already been a significant price rise for many as part of the establishment of this arrangement as OCLC adjusted their prices according to JISC banding for the previous renewal, so to see no ensuing national agreement accompanying this price rise led to a reassessment of OCLC subscriptions at many institutions, resulting in some institutions not renewing their subscriptions for this period, including a number of institutions who were members of the funnel. OCLCs subscription-based applications represent the main tools for contributing to the LoC NAF, so losing access to OCLC subscriptions resulted in many institutions losing the ability to contribute records, or not acquiring the standard tools for contribution, in the case of new joiners during this period.

I'd already enquired regarding a possible solution to this, since some institutions looking to join the funnel had not had OCLC subscriptions, and had been hoping for the JISC agreement to provide more attractively priced access, and I'd been told that alternative arrangements were possible. However, when the need for alternative access became a reality, I enquired further and discovered that the standard approach for alternative provision was for authority records and amendments by those without OCLC access was to have their contributions loaded by other institutions in the funnel.

Looking to the sky

We felt this to be a less than desirable solution, so decided to seek alternative approaches.

We considered various solutions, but the most obvious choice was the other primary application, Innovative's Skyriver, which was already established as an alternative tool for contribution to the NAF.

We'd already initiated the process for establishing access via Skyriver for NACO locally at the University of Liverpool, but before we fully established access we switched systems from Innovative's Sierra to Ex Libris's Alma, and by doing so lost access to Skyriver. However, Ex Libris and Innovative were both part of the Clarivate group, the Funnel was populated by Alma users, and I'd already attended a webinar indicating that Skyriver were looking to expand to be a more generic tool beyond Innovative products.

Negotiation and contract

Amy Staniforth from Aberystwyth enquired on the PCC list about Skyriver NACO experience which received little indication of widespread usage, and Helen Garner from Sheffield Hallam University and myself contacted Innovative, leading to an online meeting in which we discussed setting up a funnel-based access, which we discovered was not a widely established practice (only used by a funnel in Brazil as I recall?). We discussed attempting a more integrated approach involving Alma, which all those involved in the meeting used, and negotiated trial access to establish effective funnel management overall. There was some delay waiting for the contract, further delay while the various institutions spent time considering the contract (which I admit we at the University of Liverpool spent particularly long ruminating over) but eventually most of the funnel signed up.

Still looking to the sky

Not all the funnel were planning on signing up to Skyriver, however, some retaining OCLC subscriptions and being happy using their utilities for NACO access. This resulted in further delay, since Skyriver were of the understanding that all funnel institutions were going to sign up, and there were more institutions joining the funnel, some wanting access Skyriver, so there was further delay regarding sign up and co-ordination for training.

However, we've now got to a point where a date is set, and that date is quite soon (September 20th) which is just a little too late to report on for this conference, but nonetheless it's good to actually have the next step in place.

Furthermore, the University of Liverpool is more fully engaged with access via Skyriver, having had problems with the online version of OCLCs Connexion. We've been advised to use other utilities by OCLC, since online Connexion is being phased out, but since we've already planned on trying out Skyriver, we decided the timing was right to try out Skyriver as our primary tool for NACO access.

The funnel continues to expand

Meanwhile, while we've been resolving access issues, the funnel has continued to expand. Solent University are training and joining in the January 2023 intake, Open University are joining with a member who was already trained, now just awaiting Skyriver access and more staff to be trained in the September wave this year, where they will be joined by Durham University, who are also joining. Also hoping to join are University College London.

Race to train

Organizing the training for September was something of a challenge this year, however. On 7 July, there was an email on the PCC mailing list indicating that this next training wave would be open for both new trainees and refreshers, as a result of requests by many already trained cataloguers to have refreshers to take in recent RDA changes. When I saw it I immediately emailed all current enquirers to encourage signing up ASAP to beat the rush. Additional Open University staff signed up, as did Durham University. UCL were also enquiring about funnel membership, however, and were still considering at this unexpected crunch point, so are likely to join in the January 2024 wave of training.

Places were filled within a matter of hours, so it was something of a relief that we got everyone who was planning on undertaking training in the September wave signed up, enabling more institutions to make use of what continues to represent a great opportunity to empower and expand metadata operations for organizations, and to be effectively involved in the wider world of metadata management.

Bibliographic References

Grindley, N. Jisc and OCLC unable to continue agreement on National Metadata Sharing for 2022/23. 22 July 2022 https://libraryservices.jiscinvolve.org/wp/2022/07/no-agreement/ (accessed 4/11/2023)

PCC homepage https://www.loc.gov/aba/pcc/ (accessed 19/11/2023)

Kelleher, Martin, 2021. Establishing a National Funnel to the Library of Congress Name Authority File: Making it happen. *Catalogue & Index*, 203, pp. 5-10

Kelleher, Martin, 2022. Calling all funnellers! A recruitment drive. *Catalogue & Index*, 206, pp. 19-28

Kelleher, Martin, 2022. Launch of the UK NACO funnel : the start of something wonderful. 25th October 2022. https://www.youtube.com/watch?v=3rZRhOsxmEE

NACO homepage https://www.loc.gov/aba/pcc/naco/ (accessed 19/11/2023)

NAF homepage https://id.loc.gov/authorities/names.html (accessed 19/11/2023)

Laude, Marie. Intro to SkyRiver: MARC & Authority records for your library. 21 May 2020. https://vimeopro.com/innovativeiii/webinars/video/421230184 (accessed 19/11/2023)

Rediscovering prints at the British Library

Felicity Myrone, Mila Athayde, Victoria Morris, British Library

Background

Although still relatively little-known and under-utilised, the collections of prints and drawings at the British Library are the nation's largest, with particular strengths in 17th- to 19th-century British and European art. The vast majority of these artworks are contained within items in our 'Western' book, map, and manuscript collections. They are currently not well-described in our catalogues, and several mistaken assumptions still prevail:

1. That most of the nation's prints and drawings are, and have long been, held by the British Museum;
2. That the British Library's collections differ from those of the British Museum in being 'documentary' in focus and accordingly lacking in historical and aesthetic interest;
3. That the sum total of the British Library's art collections are the (in fact relatively small) sections which are visible and secure, namely our illuminated manuscripts, topographical views,[1] and 'Visual Arts', mainly the former India Office art collections.

Historic cataloguing practice has tended to prioritise text over image, and inventory over description. Thus, the British Library's 'legacy' metadata contains descriptions such as

> [Forty lithographs without letterpress, illustrating Bible history.]
> Berlin, [1874?]
> (http://explore.bl.uk/BLVU1:LSCOP-ALL:BLL01002058531)
>
> Etchings of Gothic Ornament. [Without letterpress.]
> Edward Buckton LAMB
> [London], [1830]
> (http://explore.bl.uk/BLVU1:LSCOP-ALL:BLL01002058531)

Lack of descriptive detail means that these catalogue entries are unlikely ever to be presented to users of our online catalogue,[2] and that beautiful and intriguing illustrations – such as those shown in figures 1 and 2 – are almost certain never to see the light of day.

As curators and cataloguers, we are reliant on phrases such as "with engravings" or "without letterpress"; the latter can provide a clue that a resource might comprise predominantly visual material. Equally useful are clues within shelfmarks: at the British Library, our founder Hans Sloane identified his prints with the prefix 'Pr', and the prefix "74/C." or "85/C." indicates large 'case books', a format of publication in which prints and drawings abound.

---

[1] "King's Topographical Collection : cataloguing and digitisation," 13 October 2020. [Online]. Available: https://www.bl.uk/projects/kings-topographical-collection-cataloguing-and-digitisation.

[2] British Library, "Explore the British Library," [Online]. Available: http://explore.bl.uk/

It is worth noting that there is a valid reason for a library to describe a resource as a single, illustrated volume, as opposed to a collection of prints – the volume can *only* be consulted as a volume; we're not going to unbind it (or worse still, cut the pictures out) to allow different users to view different illustrations at the same time. In simple terms: one thing merits one catalogue record.

But if this extreme brevity of cataloguing hides resources not only from the public, but from our own curators, and, moreover, if mass digitisation removes the barrier to concurrent use, then surely it is time to revisit our cataloguing and describe prints individually, and in detail?

Figure 1. Virtue preceded by a fool and followed by Glory. From 'Achillis Bocchii Bonon'. Found in the British Library at 636.g.8

Figure 2. Six mythical sea creatures. From 'Libellus varia genera piscium complectens'. Found in the British Library at. C.175.m.32.(20.)

Future goals

Felicity Myrone (Lead Curator, Western Prints and Drawings) has recently received a Getty Paper Project Publications grant to write the first handbook to the British Library's prints and drawings. The ambitious aim is for this handbook to be a ground-breaking point of entry to an under-explored collection, the first guide to the very extensive and historically significant collections of prints and drawings – but how can we write a handbook if even the curators don't know what we have?

In order to explore at least a section of our collections at scale, and at speed, we have developed a new spreadsheet-based method for cataloguing prints bound into volumes, working in collaboration with the British Museum to derive and covert their metadata. This relies heavily on two facts: firstly, that the British Museum *have* completed cataloguing of their prints and drawings;[3] secondly, that they describe prints as individual art objects, rather than components of a series or bound volume.

[3] British Museum. Department of Prints and Drawings, "Two million prints : online for all," 2022. [Online]. Available: https://www.britishmuseum.org/exhibitions/two-million-prints-online-all.

Our two institutions' shared history[4] – and our Western European bias – means that we can expect to find significant overlap between our two collections. Works were often acquired by both the Print Room (which remains part of the British Museum) and the Department of Printed Books (later to become the British Library) but disbound in the former and held intact in the latter. The challenge, to mis-quote Donald Rumsfeld, is to find the unknown knowns – the things that we don't know that we have, but that the British Museum knows about. One starting point for exploring collection commonality has been to search by series title. As an example, the word 'gravures' in the title of *Recueil de gravures au trait, a l'eua forte, et ombrees* suggests that it may be of interest for our project. The legacy metadata for this resource within the British Library catalogue leaves much to be desired:

```
=LDR nam a2200169uu 4500
=001 012807088
=003 Uk
=005 20041123123949.0
=008 040420s1809####fr#####||#####000#||fre##
=040 ## $aUk$cUk
=100 1# $aLebrun,$cM.
=245 10 $aRecueil de gravures au trait, a l'eua forte, et ombrees /$cM. Lebrun.
=260 ## $aParis :$bDidot Jeune,$c1809.
=300 ## $a2 v.
=336 ## $atext$2rdacontent
=337 ## $aunmediated$2rdamedia
=338 ## $avolume$2rdacarrier
=852 41 $aBritish Library$bDSC$jW3/7089
=852 41 $aBritish Library$bDSC$jW3/7090
```

MARC21 source data for record accessed via http://explore.bl.uk/BLVU1:LSCOP-ALL:BLL01012807088, 2023-09-21

Searching the British Museum catalogue[5] for the title *Recueil de gravures au trait* yields a family of 40 prints. Immediately we have a checklist for prints that we might find within our two volumes. We can then compare the prints that appear within our copy of the series with those listed in the British Museum metadata, rather than having to catalogue from scratch. In essence, this is copy-cataloguing as has long been carried out; the difference being that we are sourcing metadata from a museum, rather than another library.

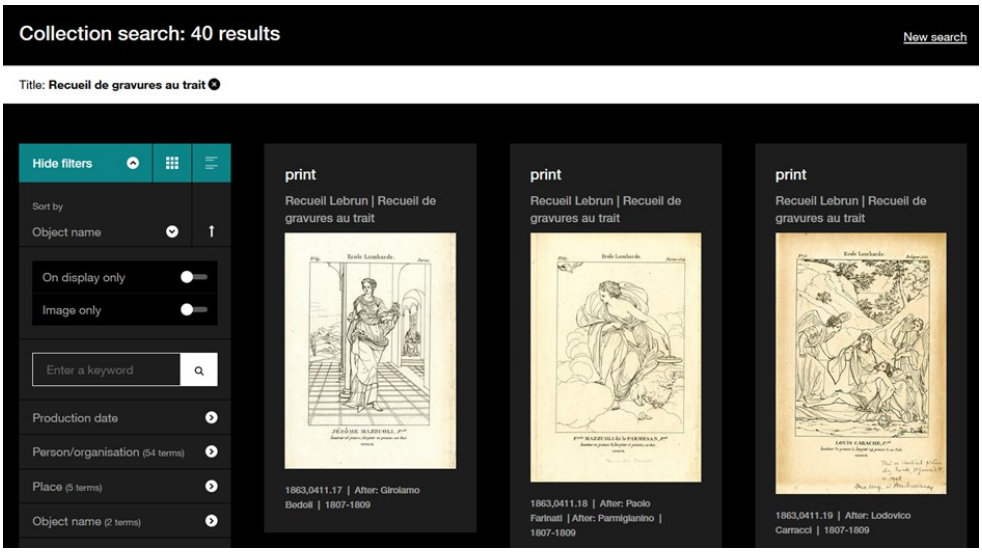


Figure 3. Screenshot from British Museum catalogue, accessed 2023-09-21

[4] F. Myrone, "Prints and drawings at the British Museum and British Library," 2017. [Online]. Available: https://www.bl.uk/picturing-places/articles/prints-and-drawings-at-the-british-museum-and-british-library.
[5] British Museum, "Explore the collection," [Online]. Available: https://www.britishmuseum.org/collection.

Whenever we are able to identify copies of the same works, we upgrade the existing record(s) for the bound volume(s) held by the British Library *and* create separate records for each of the prints contained within these bound volumes. Records for individual prints are then loaded to the British Library catalogue as analytic records. Our prints are made discoverable by printmakers, artists, publishers, former owners, titles, descriptions, mediums, and subject headings for the first time.

We describe the more technical aspects of our methodology in the following section.

Methodology

Metadata for British Museum collection items can be downloaded from their catalogue in CSV (comma-separated value) format, subject to their terms of use.[6] There is a limit of 10,000 on the number of records that can be downloaded at once, which is one reason why we did not attempt to download and interrogate the entire British Museum collection of prints and drawings at once.

We wanted to achieve two things with the British Museum metadata:

1. To upgrade our existing volume-level records;
2. To create new print-level records.

Since CSV is a text-based format, it could easily be imported into Microsoft Excel. Getting British Library metadata into the same spreadsheet was achieved by far less elegant means – we opted to copy and paste records from the 'MARC view' of our online catalogue; see cell C2 of the screenshot below for an example. In part, this was a pragmatic choice as it allowed us to get started on the project without the need to train a cataloguer in the use of library software.

	A	B	C
1	BL record ID	Filename for downloaded BM records	Paste MARC record here — Process MARC record
	BLL01000858454	collections-23-07-04-07_02_58.csv	FMT BK LDR nam a22001933 4500 001 000858454 008 990309|1804|||enk || ||||eng 040 |a Uk |c Uk 24010 |a Pictorial Representations. II. Representations originally published in Book Form or as Sets of Engravings. Holbein. Totentanz 24514 |a The Dance of Death, etc. 264 1 |a London, |c 1804. 300 |c (4º) 336 |a text |2 rdacontent 337 |a unmediated |2 rdamedia 338 |a volume |2 rdacarrier 340 |m 4to |2 rdabf 500 |a In a contemporary English gold-tooled painted calf binding. 720 |a DANCE OF DEATH |e main entry 7870 |t The Dance of Death; painted by H. Holbein, and engraved by W. Hollar. [With introductory essays by Francis Douce.] |w (Uk)MP1.0000905523 |d London: John Harding, 1804. pp. 70: plates; port. 12º. LKR |a PAR |b 000858453 |r 787 |m The Dance of Death, etc. |n The Dance of Death; painted by H. Holbein, and engraved by W. Hollar. [With introductory essays by Francis Douce.] 85241 |a British Library |b HMNTS |j C.152.f.12. 866 0 |z [Another issue.] SYS 000858454

Figure 4. Spreadsheet showing input of British Library MARC record

[6] British Museum, "Copyright and permissions," [Online]. Available: https://www.britishmuseum.org/terms-use/copyright-and-permissions.

Use of Excel dictated that further data processing was carried out using VBA (Visual Basic for Applications), the scripting language which comes built in to the Microsoft Office suite. To achieve our two stated aims, we wrote two separate scripts:

Volume-level records

The first script breaks up the British Library MARC record so that each field appears in a separate column. This allows the cataloguer to make changes to individual metadata elements.

For those interested, we include a simple version of this script in appendix 1.

Print-level records

The second script creates records for prints. For a given file of British Museum records, it looks for occurrences of the filename in column B in the previous screenshot, and creates a set of print records for each occurrence.

In essence, this is:

new print-level records = British Museum file x number of volumes

If we have multiple copies (possibly different manifestations) of the same series of prints, we can create multiple sets of print-level records at once. In the case of series such as Holbein's *Dance of death* – 19 versions found so far! – this can be a real time saving.

For the most part, the work of this second script is to pull metadata from a specific column of the British Museum metadata, and move it to the column corresponding to the appropriate MARC 21 field. Thus in Figure 5, metadata from the 'Title' column is transferred to the 'Print title [245]' column in Figure 6; similarly, metadata from the 'Producer name' column is transferred to the 'Printmaker' column.

Some metadata is also transferred from the volume-level record. For example, ownership and custodial history information (MARC field 561) for the volume is likely to apply to the prints as well. If (as is commonly the case) publication and printing information are not given on the print, these are inferred from the volume.

Object type	Museum number	Title	Producer name
print; book-illustration	No: 1871,0812.4556	Zelome Sultane (Object)	Print made by: Melchior Lorck
print; book-illustration	No: 1871,0812.4555	Verhenas Sultane (Object)	Print made by: Melchior Lorck
print	No: 1983,0127.13	Surgi fortun (Object)	Print made by: Philips Galle; After: Ni
print; book-illustration	No: 1871,0812.4557	Ruziae Soldane (Object)	Print made by: Melchior Lorck
print; book-illustration	No: 1871,0812.4637	Raheme Soltane (Object)	Print made by: Melchior Lorck
print; book-illustration	No: 1929,0416.58	Raheme Soltane (Object)	Print made by: Melchior Lorck
print; book; title-page	No: 1904,0206.107.1-121	Deß Weitberühmbten... Herrn Melchior Lor	Print made by: Melchior Lorck; Publis
print	No: 1845,0809.584	Albrecht Dürer (Object)	Print made by: Johann Gottlieb Preste
print	No: 1895,0122.782	Albrecht Dürer (Object)	Print made by: Melchior Lorck; After:
print	No: E,2.383	Albrecht Dürer (Object)	Print made by: Melchior Lorck; After:
print; book-illustration	No: 1871,0812.4558	Achada Soltane (Object)	Print made by: Melchior Lorck
print; broadside	No: E,7.253	[Sibilla Tiburtina auch Albina genandt ...] Ee	Print made by: Melchior Lorck; Publis

Figure 5. Input British Museum data

Shelfmark	Date for 008 [008]	Printmaker [100]	NACO identifier for printmaker [100]	Print title [245]
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	[Sibilla Tiburtina auch Albina genandt ...] Eer und lob ein
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	Achada Soltane
146.i.10.(22.)	s1570####	Prestel, Johann Gottlieb, 1739-1808	nb2009007182	Albrecht Dürer
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	Albrecht Dürer
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	Albrecht Dürer
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	Deß Weitberühmbten... Herrn Melchior Lorichs... Wolger
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	Raheme Soltane
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	Raheme Soltane
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	Ruziae Soldane
146.i.10.(22.)	s1570####	Galle, Philippe, 1537-1612	n86862363	Surgi fortun
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	Verhenas Sultane
146.i.10.(22.)	s1570####	Lorichs, Melchior, 1527-approximately 1590	n91091570	Zelome Sultane

Figure 6. Processed British Museum data

In addition to basic data cleaning, our script attempted to reconcile terms used in the British Museum records against authority files preferred by the British Library. The following types of term were aligned:

- British Museum relationship terms (After/Ascribed to/Book bound by/…) were aligned with MARC 21 relators;[7]
- Object types (book of prints/book-illustration/engraving/photography/…) were aligned with FAST form/genre headings;
- Subject terms were aligned with FAST subject headings;
- Names were aligned with the LC/NACO authority file[8] and ISNI.[9]

In practice, this meant keeping several lookup tables containing British Museum terms and our preferred equivalents. These could be accessed using a VLOOKUP formula. For those interested, an implementation of a VLOOKUP formula within VBA is included within appendix 2.

A certain amount of manual effort was inevitably required to build the lookup tables. However, we were able to exploit the structure of British Museum URIs to automate as far as possible. For example, we found that British Museum name authority records had URIs of the form https://www.britishmuseum.org/collection/term/BIOG followed by an integer – for example https://www.britishmuseum.org/collection/term/BIOG56009 for Sir Hans Sloane. We could therefore systematically query these records using the Python 'requests' module, extract relevant information such as name and life dates, and programmatically search the FAST and ISNI APIs. If found within the ISNI database, the ISNI record would often contain the NACO identifier, which would then allow us to retrieve the authorised form of the name from the id.loc.gov website.

[7] Library of Congress, "MARC code list for relators," [Online]. Available: https://www.loc.gov/marc/relators/relaterm.html.
[8] Library of Congress / Program for Cooperative Cataloging, "NACO : Name Authority Cooperative Program," [Online]. Available: https://www.loc.gov/aba/pcc/naco/.
[9] ISNI International Agency, [Online]. Available: https://isni.org/.

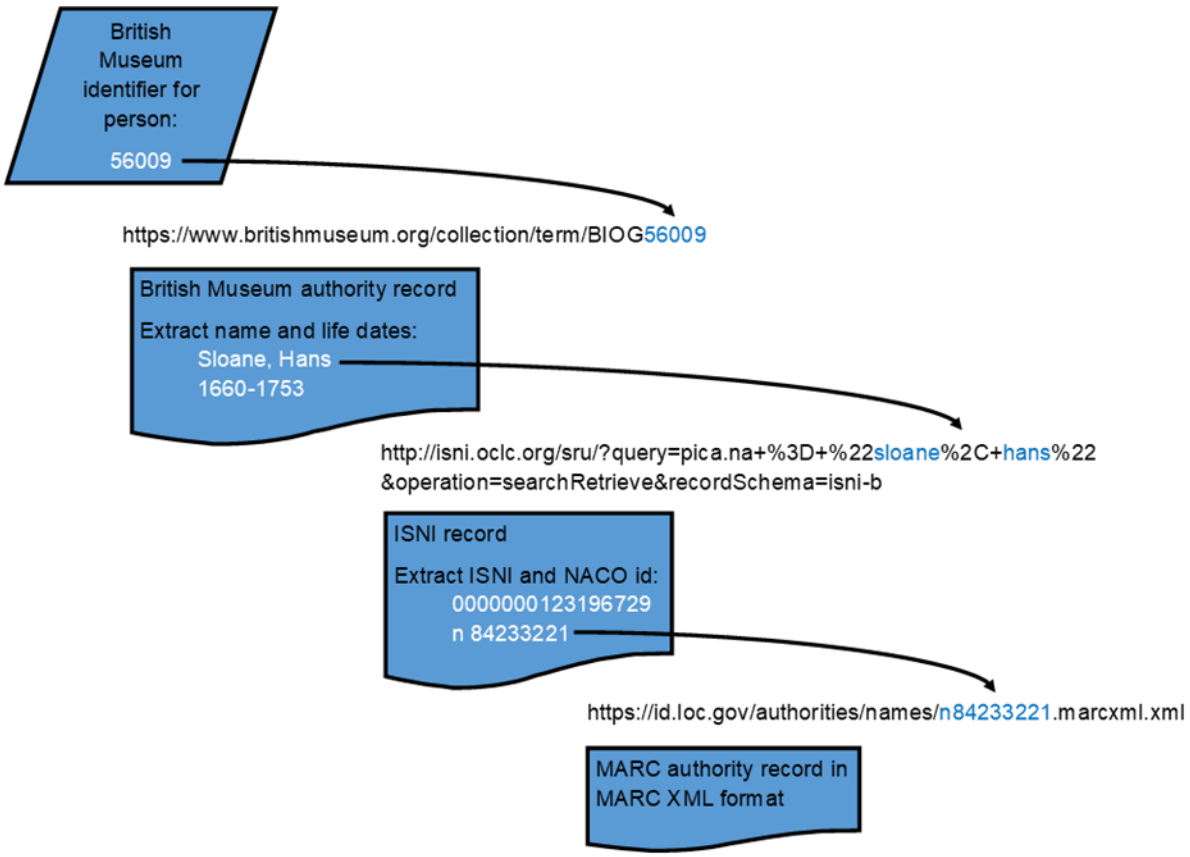


Figure 7. Flowchart for finding ISNI and NACO identifiers for names present in the British Museum catalogue. Note that the more advanced ISNI search API, which permits inclusion of life dates in the search, is only available to ISNI members. See https://isni.org/page/technical-documentation/ for technical documentation on the ISNI API.

Outcomes

To date we have been able to create records for over 350 volumes and 22,000 prints using this method. Aside from the obvious benefits of being able to open up more of our collection to new audiences, by comparing the collections of the British Museum and British Library in this way we have been able to fill in some significant gaps in our knowledge of print collection history. For example, the fact that the British Library prints are held within bound volumes has enabled us to confirm that specific prints belong to particular series, and to better identify themes depicted across these series.

One example would be the British Museum print K,68.42, depicting a female bust, divided down the centre into two sides representing day and night. Curators' comments state that, "[t]his enigmatic print was registered as part of a series of illustrations to the 1692 edition of Diogenes Laertius. This may have been a mistake."[10] We have been able to confirm that this was *not* a mistake, and that the print does indeed belong to Diogenes Laërtius' work *De vitis, dogmatibus et apophthegmatibus clarorum philosophorum*, notwithstanding that it is executed in a markedly different style from other prints within that volume.

[10] British Museum, "Catalogue entry K,68.42," [Online]. Available: https://www.britishmuseum.org/collection/object/P_K-68-42. [Accessed 2023 09 26].

Potential developments

In addition to continuing with this project, we hope to extend the methodology to other work within the British Library – and perhaps to other institutions – with anticipated benefits to the other institutions involved, as well as the wider 'print community'. A shared effort to tackle our print collections could enable a survey of which prints survive in libraries, and why, and could constitute the beginnings of a union catalogue of prints. This would be ambitious, but the benefits would be shared labour and knowledge, with the goal of revealing the nation's (or world's!) currently-hidden art collections. We'd be delighted to hear from anyone interested in pursuing this further.

Appendix 1 : VBA code to split up a MARC record into columns

This script will take a MARC record where each field appears on a separate line, and where subfields are delimited with the pipe character '|', and will split out 001, 245 $a, 245 $b and 246 $c into separate columns.

```
Function GetSubfield(sF As String, sC As String)
    'sF is the input field
    'sC is the subfield code that we are looking for
    GetSubfield = ""
    'The subfield delimiter is |
    If InStr(1, sF, "|" & sC) > 0 Then
        'Get the first subfield with the code sC
        sF = Right(sF, Len(sF) - InStr(1, sF, "|" & sC) - 1)
        'Remove any subsequent subfields
        If InStr(1, sF, "|") > 0 Then
            sF = Trim(Left(sF, InStr(1, sF, "|") - 1))
        End If
        'Trim punctuation from the end
        Do While InStr(1, ",:.; [({/\", Right(sF, 1), vbTextCompare) > 0
            sF = Trim(Left(sF, Len(sF) - 1))
        Loop
        'Trim punctuation from the start
        Do While InStr(1, ",:.; ])}/\", Left(sF, 1), vbTextCompare) > 0
            sF = Trim(Right(sF, Len(sF) - 1))
        Loop
        'Trim white space
        GetSubfield = Trim(sF)
    End If
End Function

Sub Process_MARC()

    Dim sMARC As String
    sMARC = ActiveCell.Value

    Dim aFields() As String
    Dim vField As Variant
    aFields = Split(sMARC, vbLf)

    For Each vField In aFields
        If InStr(1, vField, "001") = 1 Then
            'Record ID
            ActiveCell.Offset(0, 1).Value = "'" & Trim(Mid(vField, 5, 9))
        End If
        If InStr(1, vField, "245") = 1 Then
            'Title
            ActiveCell.Offset(0, 2).Value = GetSubfield(Trim(vField), "a")
            'Subtitle
            ActiveCell.Offset(0, 3).Value = GetSubfield(Trim(vField), "b")
            'Statement of responsibility
            ActiveCell.Offset(0, 4).Value = GetSubfield(Trim(vField), "c")
        End If

    Next vField
End Sub
```

Appendix 2 : code snippet to use ‘vlookup’ within VBA

This code assumes that the active (selected) cell contains a list of uncontrolled subject terms, separated with the delimiter ; (semi-colon). A list of uncontrolled subject terms and their authorised equivalents are stored in columns A and B of a worksheet called “Subjects”.

The ‘for each’ loop will iterate through each subject term, search for the uncontrolled term in column A, and, if found, return the corresponding value from column B.

At each iteration, Err.Clear clears any error information. After the lookup has been attempted, we test for error information (If Err.Number = 0 …); if there is no error information, then the lookup has succeeeded, and the variable sLookupValue will contain the authorised term.

```
'Subjects

For Each vVar In Split(ActiveCell.Value, ";")
    sTemp = Trim(vVar)
    sLookupValue = ""
    If sTemp <> "" Then
        Err.Clear
        sLookupValue = Application.VLookup(LCase(sTemp), _
                  Sheets("Subjects").Range("A:B"), 2, False)
        If Err.Number = 0 Then
            'Subject term has been mapped
            'Do something with sLookupValue ...
        Else
            'Subject term is not recognised
        End If
    End If
Next vVar
```

Community forward: developing an open and free cataloguing standard for rare materials

kalan Knudson Davis, Special Collections Metadata Librarian, University of Minnesota, **Jessica Grzegorski**, Head of Cataloging, Newberry Library, **Elizabeth Hobart**, Interim Head of Cataloging and Metadata Services, Pennsylvania State University, **A. Tims**, Head, Cataloging and Description, Lilly Library, Indiana University, Bloomington

Abstract

Descriptive Cataloging of Rare Materials (RDA Edition) (DCRMR) is a rare materials cataloguing standard aligned with the official RDA Toolkit. DCRMR is informed by core principles of community and sustainability while employing open-source publication models and infrastructure. The RBMS RDA Editorial Group, composed of rare materials cataloguers actively working in the field, is responsible for developing and maintaining DCRMR. This article discusses predecessor rare materials cataloguing standards that led to the development of DCRMR, the principles and constraints that shaped DCRMR from its initial inception to eventual release, the method and technical tools used to achieve the RBMS RDA Editorial Group's outcomes, and future directions for development.

[NOTE: The original version of this article was published in Library Resources & Technical Services *(v. 67, no. 1 (2023)),* https://doi.org/10.5860/lrts.67n1.26*, a peer-reviewed, open-access journal published by the American Library Association. In this version of the article, the authors have internationalised the spelling and have made minor updates to the content to reflect developments in Descriptive Cataloging of Rare Materials (RDA Edition) since the publication of the original article.]*

Introduction

Descriptive Cataloging of Rare Materials (RDA Edition) (DCRMR) is a rare materials cataloguing standard that aligns with Resource Description and Access (RDA).[1] The initial DCRMR release in February 2022 rewrote and restructured *Descriptive Cataloging of Rare Materials (Books)* (DCRM(B)) instructions to complement the RDA Toolkit.[2] DCRMR, unlike DCRM(B), is an integrating resource, published as a website, with updates framed as releases instead of revisions. While the first iteration of DCRMR contains instructions for cataloguing rare books only, future releases will incorporate instructions for other formats.

DCRMR centres community. It was created, and is maintained, by the rare materials cataloguing community for the rare materials cataloguing community. Volunteers from an array of institutions in multiple countries have contributed labour and knowledge to the standard and its growth, both by serving on the RBMS RDA Editorial Group and its predecessor groups and by providing feedback at different junctures in the editorial process. DCRMR is officially published by the Bibliographic Standards Committee (BSC) of the Rare Books and Manuscripts Section (RBMS) of the Association of College and Research Libraries (ACRL), a division of the American Library Association (ALA). However, it was created by the international rare materials cataloguing community.

DCRMR also centres sustainability. The BSC decided to create a stand-alone, integrating manual for RDA-aligned rare materials cataloguing during the ALA Annual Conference in June 2019. The RBMS RDA Editorial Group chose to create it as a website using a GitHub repository at the ALA Midwinter Conference in January 2020. Less than two months later, the COVID-19 pandemic caused massive personal and professional disruption. As our institutions closed, many of us were forced into ad hoc work-from-home situations where home and the office collided and, at times, conflicted. Then, in May 2020, the murder of George Floyd sparked a global movement toward justice-oriented community building, which became part of the warp and weft of creating DCRMR, as much of the technical infrastructure of DCRMR was built by an Editorial Group member who resides in the Powderhorn Park neighbourhood of Minneapolis. Over the past several years, it has become clear that in order to be sustainable, DCRMR must rely on communal efforts, not individuals, to allow for people to step back, step down, and maintain their own well-being. Towards that end, succession planning, open sharing of knowledge, and extensive documentation have been integrated into DCRMR's workflows and planning.

DCRMR centres open-source infrastructure. Built on a zero-dollar budget, it is available openly and freely on the web, aligning our cataloguing standards with our professional values of transparency, accessibility, and

equity.[3] The Editorial Group utilized freely-available, well-established technical tools and software like Python, Ruby, GitHub, and Notepad++ to create DCRMR, and the text is published under a Creative Commons license that allows others to adapt the text to their own needs or use the code base to develop their own infrastructure. Choosing an open-source model allows people to use and build on the work of the Editorial Group, just as the Editorial Group has used and built on the work of others.

DCRMR is one contribution in the overarching history of rare materials cataloguing standards development.

Background and literature review

Raw Materials Cataloguing Standards

Rare materials have distinct cataloguing needs, including describing individual issues and states and distinguishing individual copies of a manifestation. In an article on the development of rare book cataloguing practices, Beth M. Russell highlights the "constant tension between descriptive bibliography and library cataloging."[4] Russell notes the fundamental differences between "mainstream" cataloguing and rare materials cataloguing, chiefly the artefactual nature of rare materials due to their method of construction.[5] She highlights the philosophical differences between various cataloguing codes of the past while emphasizing the importance of transcription, transposition, format and collation, and the robust nature of rare materials notes, access points, and copy-specific information.[6] Similarly, Juliet McLaren and Jane Gillis compile a history and development of rare serials cataloguing rules, emphasizing that rare serials "cannot be identified without careful transcription of their sometimes unique extended titles, their imprints (where present), and detailed notes."[7] Their analysis of the descriptive needs for rare serials walks through each area of description from International Standard Bibliographic Description (ISBD), detailing why previous rules were inadequate for identification and ending with a discussion of cataloguing early reprints, republications, and access points.[8]

Throughout the history of rare materials cataloguing, cataloguers have sought to reconcile these needs with the strictures of cataloguing codes. Russell's article discusses the reconciliation process between bibliography and G. Thomas Tanselle's argument for "mutual understanding between bibliographers and catalogers" in the 1970s and the nature of recording physical facts.[9] Russell then describes the evolution of rare materials cataloguing through various past codes.[10] Mary Burns continues this discussion, detailing the evolution of rare materials cataloguing standards by summarizing the development history of *Bibliographic Description of Rare Books* (BDRB), *Descriptive Cataloging of Rare Books* (DCRB), and the various task forces leading to the development of the RBMS Policy Statements (RBMS PS), previously slated for incorporation into the RDA Toolkit.[11] Burns, in her two-part article "RDA and Rare Books Cataloging," compares the cataloguing outputs of three bibliographic records created for the same book following the stipulations of DCRM(B), the *BIBCO Standard Record (BSR) RDA Metadata Application Profile* with rare materials provisions, and the original RDA Toolkit (2013) with its exceptions for early printed resources.[12] Burns notes that, even with the provisions and expectations, "there are description and transcription issues that rare materials catalogers need to address that RDA, a general cataloging standard, does not," suggesting that the discrepancy between rare materials standards and general standards remains.[13]

In 2007, the BSC published DCRM(B) in collaboration with the Library of Congress.[14] DCRM(B) was the first in the suite of *Descriptive Cataloging of Rare Materials* (DCRM) manuals. Meanwhile, the RDA Steering Committee (RSC) (formerly the Joint Steering Committee for Development of RDA) began developing RDA to replace the second edition of AACR2 as part of its strategic plan (2005-2009).[15] The editors of DCRM(B) considered postponing work on the manual until the publication of RDA but elected to proceed, "given the progress already made on DCRM(B) and the considerable investment to date of time, labour, and money."[16] Subsequently, the BSC published five DCRM manuals covering additional formats: serials (2008), graphics (2013), cartographic (2016), manuscripts (2016), and music (2016).

After the publication of RDA in 2010, rare materials cataloguers quickly began to consider the future of DCRM in relation to RDA.[17] Dr Robert Maxwell and John Attig investigated issues surrounding the future development of DCRM following the adoption of RDA, including the relationship between the standards, terminology used within the DCRM text, the organization (i.e., structure and arrangement) of the standard, descriptive aspects not traditionally covered by DCRM, DCRM's relationship to *International Standard Bibliographic Description*

for Older Monographic Publications (Antiquarian) (ISBD(A)), and broader policy related to the application of DCRM(B).[18] Their discussion paper also outlined differences between *Anglo-American Cataloguing Rules*, 2nd ed. (AACR2) and the original RDA Toolkit that are relevant to the revisions of DCRM, including differences in terminology, sources of information and use of brackets, transcription practices, use of abbreviations, categorisation of resources using RDA elements (e.g., media type, carrier type, etc.), recording terms from controlled vocabularies, and the formulation of access points for manifestations and items.[19] In the years between Maxwell and Attig's discussion paper and the initial development of DCRMR, this report has served as a touchstone for the intervening task forces and editorial groups.

Todd Fell and Francis Lapka posed the possibility of an international standard for rare materials cataloguing.[20] They outlined several requirements for this standard: an extension of a standard for general cataloguing that acknowledges the needs of the specialist community, an international governing body with translations for use in diverse communities, embraces the prevailing international models for bibliographic description, is open and reusable, acknowledges the centrality of transcription in rare materials cataloguing, integrates with the current data landscape, and is responsive to user needs.[21] Although this article did not address whether there should be a common standard for rare materials cataloguing, it did offer one possible path forward for this work.[22]

The BSC formed the DCRM-RDA Task Force (2011-2012), which recommended revising DCRM(B) to align it with RDA.[23] In 2012, the BSC formed the DCRM(B) for RDA Revision Group to complete this work.[24] The Program for Cooperative Cataloging (PCC) released the first iteration of the BIBCO Standard Record (BSR) on 1 January 2013.[25] The BSR includes DCRM-aligned provisions for cataloguing rare materials developed in collaboration with the PCC Task Force for BSR for Rare Materials Based on RDA. On 22 April 2013, the BSC issued a statement on the relationship between DCRM and RDA, stating that the BSC is "neutral … neither encouraging nor discouraging agencies regarding implementation of RDA-acceptable DCRM records."[27] The statement provided interim guidance to cataloguers using DCRM until an RDA-aligned version of DCRM could be published. For most rare materials formats, cataloguers could choose either to follow the appropriate AACR2-based DCRM manual for description in conjunction with RDA for constructing access points or to create RDA records using the rare materials provisions in the BIBCO Standard Record.

At the ALA Annual Conference in 2013, the BSC expanded the charge of the DCRM(B) for RDA Revision Group to create RDA-aligned guidelines for all formats in the DCRM suite and renamed the group the DCRM for RDA Revision Group.[28] At the next ALA Annual Conference in 2014, the Revision Group recommended authoring a set of policy statements for rare materials to accompany RDA instead of rewriting the DCRM suite. In response, the ACRL/RBMS Descriptive Cataloging of Rare Materials Task Force (2014-2017), an independent RBMS task force under the aegis of the BSC, was established to complete this project.[29] In 2016, the task force formally named its guidelines the RBMS Policy Statements (RBMS PS) in alignment with the naming conventions of other RDA policy statements.[30] In 2017, the Descriptive Cataloging of Rare Materials Task Force submitted an initial draft of the RBMS PS and disbanded.[31] The BSC absorbed the editorial work on the policy statements, but much of the work was put on hold while the RSC revised the RDA Toolkit in response to the RDA Restructure and Redesign (3R) Project.[32]

Uses, benefits, and workflows of Git and GitHub in library science

Even a cursory glance into library science literature will illuminate the many and varied uses, benefits, and workflows of GitHub. Robin Camille Davis lists examples of the use of GitHub in a library context, including developing and sharing code or datasets, digital archives, or writing entire books, highlighting that "GitHub has become a site for academic transparency" and calling Git a "librarian's dream tool."[33]

In addition to transparency, Davis discusses the following benefits of GitHub: version control, ease in creating documentation, and social networking.[34] Prayudi Utomo and Falahah describe the benefits of developing a serverless website hosted using GitHub Pages, including increased productivity, ease of website management and configuration, and reduced effort for code review while implementing new services.[35] In this instance, the authors chose GitHub Pages as their Content Delivery Network (CDN) because of its version tracking, robust collaboration support, and free static website hosting.[36] Yasset Perez-Riverol, et al. remark

that GitHub eases "sharing programming tasks between different remote contributors," while the version control system provides transparency in the development process and the inbuilt social features support "peer review, commenting, and discussion."[37]

Keith Engwall and Mitchell Roe outline a typical Git workflow describing a main branch and the creation and merging of development branches onto the main branch.[38] Their six-step workflow used in a web development model includes: creating a discussion issue for a proposed change, creating a development branch for the proposed change, editing code and testing the development branch until the change is complete, undergoing a development code review process, merging the development branch into the main branch, and pushing the changes to a production web server.[39] Because of the numerous benefits of GitHub for collaborative workflows and projects, it is central to the development of DCRMR.

The move toward DCMR

In August 2018, the BSC formed a subgroup to finalize the draft of the RBMS PS for publication in the RDA Toolkit.[40] In April 2019, the RSC completed the 3R Project and released a stable English-language version of the Toolkit. However, the substantial changes to the Toolkit meant that the RBMS PS could not be used in their current form.

Following discussions at the ALA Annual Conference in June 2019, during which the rare materials cataloguing community expressed a desire for a stand-alone manual, the RBMS Policy Statements Editorial Group decided to rewrite the DCRM suite as a single RDA-aligned integrating resource and write lightweight policy statements to link from the RDA Toolkit to the revised DCRM.[41] To reflect this change in scope, the group was renamed the RBMS RDA Editorial Group. In February 2020, the new manual was officially named Descriptive Cataloging of Rare Materials (RDA Edition) (DCRMR).[42]

The RBMS RDA Editorial Group consists of 10-14 members. One or two members serve as chief editors and are responsible for Editorial Group planning, finalizing editorial decisions, maintaining high-level consistency across the text, and liaising with external groups as appropriate. In addition, at least two members serve as keepers of the text (also called keepers), who are responsible for developing and maintaining the GitHub deployment and maintaining the canonical version of the text. All group members play an editorial role by participating in the drafting and revision of text and in the collaborative decision-making process.

Principles and constraints

DCRMR was conceived and built to meet the need of the rare materials cataloguing community for a stand-alone manual, using language that will be familiar to cataloguers and clear cataloguing instructions with citation numbers to assist in citing a particular instruction. To support practical applications of the DCRMR instructions, all examples represent real-world objects and descriptions to better reflect cataloguing in practice.

While earlier DCRM manuals were published as monographs, DCRMR is an integrating resource, which allows the text to be responsive to changes in RDA. In addition, the manual is published online as an open-access resource, ensuring broad availability. DCRMR is available to all interested users for free, both via the internet and via a downloadable PDF, allowing users to print the document if they wish. DCRMR is licensed with a Creative Commons Attribution NonCommercial ShareAlike ((CC) BY-NC-SA) license, allowing others to adapt the text to their local needs.[43]

The Editorial Group has sought to maintain transparency throughout the process. Because the text is hosted on GitHub, users are able to submit issues (a discussion thread on problems encountered or future developments) and read discussions of those issues. GitHub is a version control system, allowing users to see how the text has changed over time. Finally, the Editorial Group retains earlier versions of the downloadable PDF for any users wishing to consult earlier versions of the text.

Material constraints have impacted publication. DCRMR is created and maintained by a volunteer committee. Editorial Group members receive no compensation and need to schedule around other personal and

professional commitments. The Editorial Group prepared the first iteration of DCRMR between January 2020 and July 2021. During this time, many Editorial Group members worked remotely or on hybrid schedules, allowing for extra writing time. At the same time, the global events that occurred during 2020 and 2021 took a significant toll on group members. For past DCRM manuals, the Editorial Group met in person to discuss editorial decisions. The pandemic forced the RBMS RDA Editorial Group to collaborate virtually. The group met in person for the last time during the ALA Midwinter Conference in January 2020, about seven weeks before lockdowns began in the United States. Although the group met consistently throughout the pandemic, they never expected that work would be exclusively virtual. The inability to meet in person, combined with the significant stress posed by the events of 2020 and 2021, delayed the initial publication by a year from the original timeline.

In addition, the Editorial Group created DCRMR without direct financial support. They rely instead on freely-available tools with no paid developer support. At times, this leads to problems, such as advertisements appearing as part of the Google Programmable Search Engine or minor technical difficulties.[44]

Method

The RBMS RDA Editorial Group needed to create and sustain an iterative, integrating resource that would incorporate additional DCRM manuals in the future, be responsive to changes in RDA (itself an integrating resource), and be maintained and updated by a succession of future group members. To do so, the group developed a cyclical workflow that oscillates between Google Docs and GitHub and that is buttressed by extensive documentation and facilitated by both Python scripts and human labour.

The initial text

In autumn 2019, the Editorial Group began writing what would become DCRMR. To begin, they atomized the DCRM(B) text into multiple Google documents, one RDA element per document. The Editorial Group omitted examples, textual numbering, and text about prescribed punctuation at this time because they intended to holistically review and standardize their approach to these topics. The group edited the text to bring it into alignment with RDA terminology and incorporate decisions made in the now-superseded RBMS PS. They raised smaller issues using the Google Docs comment feature and discussed larger questions through the Editorial Group listserv and during virtual meetings.

During this revision stage, the Editorial Group also made decisions on the structure of the text, which they later built into the website's architecture. To respond to the community's desire for a manual in workflow order, they decided to retain a chapter structure rooted in ISBD. Significant changes to the order of the text from DCRM(B) include:

- restructured elements related to statements of responsibility as an independent chapter. In DCRM(B), instructions related to title and statement of responsibility are both in Chapter 1, "Title and statement of responsibility area"; in DCRMR, instructions for statement of responsibility are in Chapter 2, "Statement of responsibility";

- incorporated notes into the relevant chapters. For example, in DCRM(B) all instructions for notes are found in Chapter 7, "Note Area" (7B3-7B5); in DCRMR, instructions for the element Note on title are found in Chapter 1, "Title" (1.29). DCRM(B)'s Chapter 7, "Note Area," became DCRMR's Chapter 9, "Additional notes"; and

- added placeholder chapters to hold space for the integration of additional DCRM manuals in the future. Specifically, DCRMR includes Chapter 4, "Mathematical details," for cartographic description and Chapter 7, "Numbering of serials," for serials description.

Once the Editorial Group determined the order of the text, they crafted a citation scheme. Since DCRMR is an integrating resource which will both incorporate additional instructions and respond to changes in RDA, they decided to use a four-part decimal-based citation scheme to allow for greater flexibility and extensibility. The

citation scheme is a mix of numbering that carries meaning (for example, the numbers in Chapter 3, "Edition," start with "3") and numbering that is arbitrary (for example, most element numbers started with ".2" to allow space to insert elements earlier in the chapter).[45]

Migration to Github

With the order of the text in place, the keepers could build the initial website architecture, and the Editorial Group could start migrating the text from Google Docs into the GitHub repository. During summer 2020, keepers conducted training sessions for interested Editorial Group members, demonstrating how to format the text using Markdown, a lightweight markup language, and save the resulting Markdown files to the GitHub repository. Throughout the summer and into early autumn, the keepers and group members migrated the text as it was completed and reviewed. Each of the atomized Google Documents, one RDA element per document, became the basis for the Markdown files. From this point, completed drafts of the Markdown files containing instruction text lived in the GitHub repository and could be viewed as a whole and in context on the website. Working copies were kept in Google Docs, where editing, revision, and review occurred.

Reviewing the newly migrated text also allowed the chief editors and keepers to see variations in writing style, textual formatting, and input conventions. In order to ensure uniformity across the text, the keepers developed a detailed style sheet that included instructions on how to mark up and input text.[46] Some guidelines are quite granular (for example, "Alternative rules are introduced by 'Alternative rule', formatted in bold and followed by a period. The period is not in bold.").[47] Others provided broader, more flexible instructions (for example, "Alphabetized lists preferred. However, numbered lists are sometimes appropriate to the text or necessitated by the display.").[48] Chief editors and keepers discussed decisions about style. The chief editors brought some questions, like link formatting, to the whole Editorial Group for discussion. In order to minimize future variations in the style and formatting of the text, the Editorial Group centralized the editing of Markdown files in GitHub as part of the keeper role.

Developing Cyclical workflows

Once the keepers migrated the text to GitHub, the Markdown files served as the canonical copy of DCRMR, and the Editorial Group members could read and review it holistically. The keepers export the entire text from GitHub Markdown files into Microsoft Word documents using a Python script that they developed for this purpose. This is usually done on a chapter level, although more targeted text selections are sometimes desired. The Word documents are then uploaded to Google Drive and converted to Google documents. Editorial Group members can then perform detailed, line-level textual markup on these working files and conduct associated paratextual conversations using familiar tools. Group members read, analyse, and suggest revisions to the text using the "Suggest" mode. The chief editors then review all suggested changes, rejecting undesired changes and leaving approved changes as suggestions.

This process allows the chief editors to maintain a high-level view across the text and ensure consistent decisions throughout. The keepers then implement the changes in the Markdown files in GitHub and correspond with the chief editors for clarification as necessary. For complicated edits (for example, reordering text, which affects both the text undergoing edits and any link to or citation of that text), the chief editors and keepers may utilize tracking spreadsheets and other supplemental, ad-hoc documentation in order to complete all needed changes.

Once the text is ready for feedback by a community constituency, such as the editorial group's parent body, RBMS BSC, or the international rare materials cataloguing community, the text is frozen. The reviewing community is directed to the website for the development fork, which is generated from the revised text in the GitHub repository, where they may read and navigate DCRMR as a hypertext document. This GitHub fork is an exact copy of the DCRMR repository; however, the fork's environment, which determines the content of its website, is set to the development branch of the repository, allowing the changes to the Markdown files to be reviewed in context while leaving the production website, which displays the canonical version of the text, unaltered during the review period. The Editorial Group uses Google Forms to collect feedback. During the

review period, links to the forms are added to the DCRMR website, often on the chapter level. This has the advantage of gathering all feedback in a central location.

Following the review period, the chief editors review all feedback. The keepers make small bug fixes and correct minor issues, such as typos and broken links. More substantive issues may be addressed by the chief editors or through Editorial Group discussion and then incorporated into the text; some issues are flagged for future discussion. Once the Editorial Group completes post-review edits, the editorial cycle begins again. Figure 1 illustrates the cyclical editorial workflow and the tools involved in DCRMR revision.

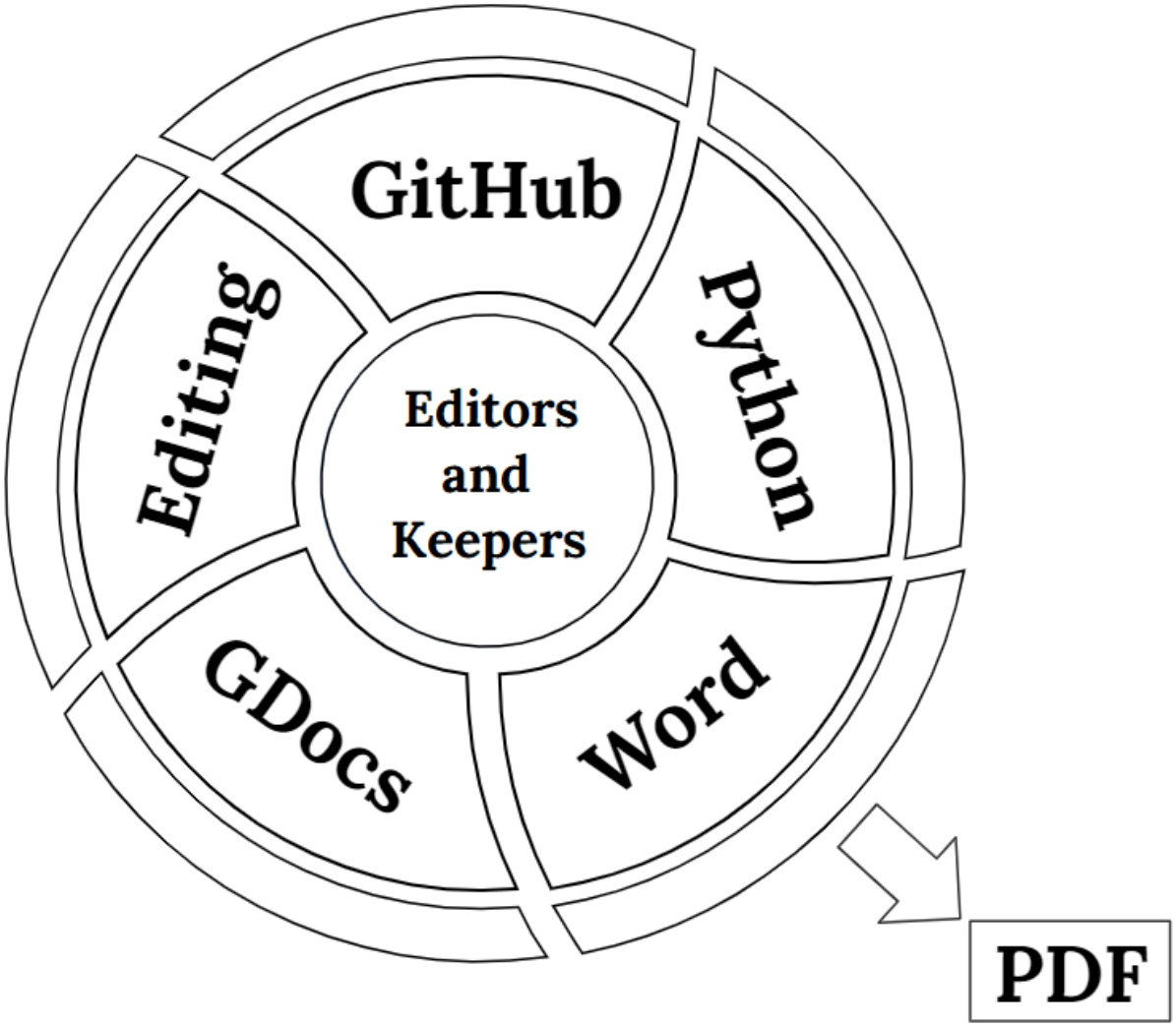


Figure 1. The cycle of tools used in DCRMR revision.

Technical tools

The technical tools chosen to build DCRMR are rooted in the same open-source and community-forward principles and limited by the same budgetary constraints that drove its initial creation. Many of the technical tools are available for free and support asynchronous web development. They facilitate the work of multiple contributors located in different geographic areas with varying levels of technical expertise. The keepers used the following tools heavily throughout the DCRMR development process.

GitHub

GitHub is the largest open-source community in the world; it contains millions of projects with a focus on growing skills and helping others by building healthy communities of contributors.[49] Discussions surrounding GitHub began when Liz Adams and Francis Lapka prepared an internal report on the various hosting options at the request of the DCRMR editors.[50] GitHub offered several advantages over other hosting options, including version control, issue tracking, public access to wiki documentation, pull requests, project planning tools for future releases, and a lightweight formatting syntax (i.e., Markdown). Finally, GitHub offered a range of scenarios for publication such as publishing as a single Markdown file (similar to the implementation of *Describing Archives: A Content Standard*) or as multiple files hosted within a repository using github.io or a custom domain.[51] Ultimately, the RBMS RDA Editorial Group decided to implement GitHub with multiple files hosted on a custom domain, publishing DCRMR on a subdomain of the RBMS website (https://bsc.rbms.info).

DCRMR's Repository

The GitHub repository contains the text of DCRMR and the codebase that powers the bsc.rbms.info website on GitHub Pages. The DCRMR repository also hosts various picture files, assets, and scripts used in creating and maintaining the website and text. The canonical, current, and approved version of the text is contained in the main branch of the repository. Revisions to DCRMR text are contained in branches and merged upon the chief editors' approval. Figure 2 illustrates GitHub branches as used in DCRMR revisions.

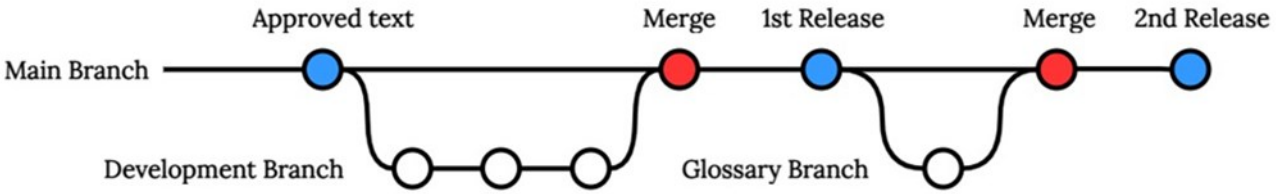


Figure 2. GitHub branches in DCRMR revisions.

Although much emphasis has been placed on the reasoning behind choosing GitHub as a home for the development of DCRMR, the keepers rely on many other free and open-source tools to ease the upkeep and ongoing maintenance of the website. Keepers working in a Windows environment must download Git for Windows, which is a free and open-source BASH emulation allowing Windows users to run Git from the command line.[52]

In the deepest recesses of DCRMR's heart is Ruby, a free and open-source programming language with an emphasis on simplicity, productivity, and elegance.[53] Ruby utilizes a standard format for distributing programs and libraries in a "gem." Jekyll is a static site generator installed using Ruby. The Jekyll gem, along with several other Ruby gems, are installed using Git Bash. The Jekyll gem takes the Markdown files containing the DCRMR text and converts them into a complete, static website. Again, the premise of Jekyll is rooted in openness and configuration simplicity with an emphasis on content.[54] Because DCRMR is a tool created by and for cataloguers, simplicity, ease of software maintenance, and freely available tools are paramount.

DCRMR uses Minimal Mistakes, a flexible two-column Jekyll theme, for creating and customizing the website's presentation. The Minimal Mistakes theme includes all the assets, html layouts, and cascading style sheets that give the website its overall look and feel.[55] The keepers carefully document customizations to DCRMR's implementation of Minimal Mistakes to ensure that users of the DCRMR website will continue to have a similar end-user experience as future Minimal Mistakes releases are tested and implemented.[56]

In addition to using all the development tools above, the keepers use Notepad++, an open-source code/text

editor, to create and edit the Markdown files in the DCRMR GitHub repository.[57]

Jupyter Notebooks & Python

The keepers developed the Python script, which is used to compile the Markdown files into a Word file, in an environment that upholds the same basic principles and tenets as DCRMR itself. Anaconda Navigator is a desktop application that manages integrated applications, packages, and environments in an open-source, user-friendly, and community-centred development platform with open documentation, describing itself not as a company but rather as a movement.[58] This resonates strongly with the underlying principles of DCRMR, which are as much about a movement toward aligning our professional values with cataloguing rules as they are about rare materials cataloguing.

The keepers created the Python script in Anaconda Navigator's Jupyter Notebooks. The script is iterative in nature and evolves over time, enabling the cyclical editorial process between GitHub and Google Docs. The keepers test the script in a branch of the DCRMR GitHub repository and, once they have sufficiently tested the improvements, it is merged into the main repository.[59] Recent scripting improvements include preserving formatting and DCRMR's structure when converting from Markdown files to Word documents utilizing pypandoc, a universal document converter, and docxcompose, a Python library for concatenating and appending Microsoft Word (.docx) files. The editable script allows for the structure of DCRMR text to change over time as sections of instructions are drafted and new Markdown files are added to the static GitHub Pages hosted website.

Google Docs

Through 2019, the Editorial Group primarily used Google Docs as a platform to craft the RBMS PS. As mentioned above, Google Docs remains an integral part of the group editing and revision process. Google documents are still utilized heavily in the DCRMR revision and review cycle, but solely as a way for the chief editors to collect feedback or the Editorial Group to further refine the text for the next DCRMR GitHub release.

Succession planning and sustainability

The long-term sustainability of DCRMR depends on not relying on any single person's technical skills, availability, or institutional memory. It will be a multi-year project to incorporate all of the formats in extant DCRM manuals, and it will require a range of skills and contributions in cataloguing knowledge and format specialties. Likewise, maintaining and updating the website for an indefinite period will also require the skills of many GitHub contributors. In order to facilitate sustainability, the Editorial Group is taking a multi-pronged approach: collaboration in key roles, active succession planning, and extensive documentation.

Membership in the RBMS RDA Editorial Group will shift over time. To facilitate changes in membership, the Editorial Group established a model of assigning co-chief editors and co-keepers. This distributes responsibility across multiple individuals; if one person is busy, the other person can usually step in to make sure the project continues to move forward and deadlines are met (deadlines can and have been moved, as well.)

Additionally, the Editorial Group established shared accounts for activities in GitHub so that access is not tied to a single individual but rather to member roles. The DCRMR repository is owned by the RBMS Bibliographic Standards Committee GitHub account, rbms-bsc, which provides access to current maintainers and contributors. The Editorial Group's GitHub account, dcrmr, owns the forked development repository, and the chief editors use it to close issues after the resolution of an editorial discussion. Keepers of the text generally complete revisions to DCRMR text, website code, and scripts by using personal GitHub accounts with commit access to the repository.

In order to ensure continuity, the Editorial Group has established staggered terms for the chief editors. Terms change every July, following the annual volunteer cycle of ACRL; the incoming chief editor is selected by the previous spring. The keepers are working to establish a similar staggered succession model, following a three-year cycle of incoming, established, and emerita positions.

The DCRMR repository contains a wiki with official documentation both for internal and external audiences.[60] The Editorial Group uses the wiki to host internal documentation on DCRMR's editorial and style guidelines, and citation scheme; instructions on website maintenance, running scripts, and setting up their computers to perform DCRMR editing via GitHub Desktop and a local environment; and templates for new DCRMR text. Other documentation for both the Editorial Group and the general public includes a DCRMR FAQ page, current and historical RBMS RDA Editorial Group membership, reported errata, and resources on succession planning and leadership transitions. In alignment with DCRMR's principles on succession planning, the wiki documentation on the Python script evolves over time as new keepers take on roles within the organization of the Editorial Group. As many cataloguers are just starting to actively build skills in Python and GitHub, the documentation helps to build confidence in successive keepers, guiding them through the steps of downloading Anaconda Navigator, installing Python packages, placing Markdown files, and running the script or creating a Ruby/Jekyll environment on their local machines for website development and testing.

Outcomes

The BSC officially published DCRMR on 2 February 2022, following a vote from the RBMS Executive Committee. At the time of this writing, DCRMR has been used to create or edit over 8,200 records in OCLC. DCRMR has generated global interest. One hundred and fifty participants from eight countries spanning three continents attended the public hearing sessions on DCRMR in December 2021. In the first year after initial publication, Google Analytics shows access from 13,768 users from 111 countries, representing six continents (see Table 1).[61]

Country	*Users*
United States	8,128
United Kingdom	1,008
Canada	734
Australia	649
Netherlands	498
Finland	411
France	310
Austria	194
South Africa	172
Germany	140
Japan	112
Philippines	101
India	80
Czechia	79
China	64
Sweden	63
Spain	58
Norway	55
Singapore	54
South Korea	51

Table 1. Number of DCRMR users from the top twenty countries, 3 February 2022 to 2 February 2023.

Future directions and development

DCRMR is currently a minimum viable product incorporating instructions for rare book cataloguing only. The first major update to DCRMR was released in November 2022, which added a glossary, added general rules for all elements, and updated two element names to reflect changes in RDA. To maintain transparency, a changelog and semantic versioning were incorporated into DCRMR, and the keepers developed maintenance tools and routines for citation numbers. In the longer term, the Editorial Group plans to incorporate instructions for the remaining five formats covered in the original DCRM suite, starting with graphics. The Editorial Group documents ongoing maintenance and future work in the DCRMR GitHub repository.[62] These tasks include updating the early letterforms and brevigraphs tables in DCRMR, incorporating additional examples, and drafting sections on pre-cataloging decisions and other topics.

Like RDA, DCRMR is an integrating resource that will be updated over time, and editorial work on the standard follows an iterative process. The RSC generally releases updates of the RDA Toolkit four times a year.[63] The Editorial Group will review the release notes after each update and make any necessary changes to DCRMR so that it remains aligned with RDA. Major changes to RDA that will require revisions to DCRMR include revisions to RDA element names or definitions and the deprecation, or phasing out, of elements over time. The Editorial Group will also respond to changes in best practices for rare materials cataloguing. Before any major updates to DCRMR, particularly before the integration of instructions for additional formats, the Editorial Group will solicit and carefully consider community feedback.

Along with future release cycles for revised and new sections of DCRMR instructions, the freely hosted infrastructure will be updated and rigorously tested as developers release new versions and patches for the various software and tools that power the website, such as Git, GitHub, Ruby, Jekyll, and Minimal Mistakes. The first minor update to DCRMR was released in May 2023 and added a Citation File Format (CFF) file to the root of the DCRMR GitHub repository's structure which allows the automatic population of a preferred citation for DCRMR as a bibliographic resource through popular citation managers, such as Zotero and Mendeley, and in various citation formats, such as APA, Chicago, etc.[64] In addition to the creation of this Citation.cff file, the keepers of the text also began minting Zenodo DOIs for GitHub releases. A Zenodo DOI increases the amount of nuance and accuracy in bibliographic citations for DCRMR as the text evolves and changes over time. For example, one DOI represents the overarching concept of DCRMR as a whole and will always resolve to the latest version.[65] Other DOIs are minted for specific DCRMR releases, such as the first major release. Each of these DOIs preserves a snapshot of how the repository and text existed in the past.[66]

Conclusion

DCRMR is a standard made by the rare materials cataloguing community for the rare materials cataloguing community. The RBMS RDA Editorial Group followed the guiding principles of accessibility, openness, and sustainability throughout the development of the standard. Because DCRMR is hosted and built with free, open-source tools, such as GitHub and Jekyll, any cataloguer may access and use the standard at no subscription cost. Under the provisions of DCRMR's Creative Commons license, cataloguers, individual institutions, and other organizations can adapt the text to their needs or use the base code in GitHub to develop other open cataloguing standards.

The Editorial Group is committed to transparency and open collaboration. Anyone interested in the development of DCRMR may read and comment on the issue threads in DCRMR's GitHub repository or consult the extensive documentation posted in the Editorial Group's wiki. The use of open, collaborative, and familiar tools such as Google Docs ensures that group members with varying levels of technical expertise can participate equally in the editorial process.

Finally, the Editorial Group selected open tools where possible and developed an editorial process with sustainability in mind. GitHub and Python, for example, are well established; thriving communities of users across many domains continue to implement these tools in a variety of applications. The use of Python scripts, in particular, automates many of the routine maintenance tasks for DCRMR, freeing time for editorial work and the ongoing development of the website. The Editorial Group's model of staggered terms for co-chief editors

and co-keepers, as well as wiki documentation on succession planning and onboarding new group members, bolsters the long-term sustainability of DCRMR.

The RBMS RDA Editorial Group, a dedicated group of volunteers who work in the field, is committed to developing and supporting a standard that is broadly useful to the rare materials cataloguing community. It takes a village to raise a cataloguing standard, but it takes a community to make it thrive. Feedback from and dialogue with peers working in rare materials cataloguing are essential to the ongoing relevance and utility of DCRMR.

References

[1]Descriptive Cataloging of Rare Materials (RDA Edition) (v2022.2.1.1), ed. RBMS RDA Editorial Group, Bibliographic Standards Committee, Rare Books and Manuscripts Section, Association of College and Research Libraries, accessed 27 May 27 2022, https://doi.org/10.5281/zenodo.7953732; RDA Steering Committee, "RDA Toolkit," accessed 27 May 2022, https://access.rdatoolkit.org.

[2]*Descriptive Cataloging of Rare Materials (Books)*, ed. Bibliographic Standards Committee, Rare Books and Manuscripts Section, Association of College and Research Libraries (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 2011), accessed 27 May 2022, https://rbms.info/files/dcrm/dcrmb/DCRMB3.pdf.

[3]Cataloging Ethics Steering Committee, *Cataloging Code of Ethics,* last modified January 2021, accessed 15 February 2024, http://hdl.handle.net/11213/16716.

[4]Beth M. Russell, "Description and Access in Rare Books Cataloging: A Historical Survey," *Cataloging & Classification Quarterly* 35, nos. 3-4 (2003): 492, https://doi.org/10.1300/J104v35n03_10.

[5]Ibid., 493.

[6]Ibid., 494-496.

[7]Juliet McLaren and Jane M. Gillis, "Is This Rule Necessary? A Discussion of New Rules for Rare Serials," *Cataloging and Classification Quarterly,* 43, no. 1 (2006): 41, https://doi.org/10.1300/J104v43n01_04.

[8]Ibid., 43-51.

[9]Russell, "Description and Access in Rare Books Cataloging," 497.

[10]Ibid., 499-514.

[11]Mary Burns, "RDA and Rare Books Cataloging, Part 1," *Library Resources and Technical Services* 62, no. 4 (2018): 160-162, https://doi.org/10.5860/lrts.62n4.160.

[12]Ibid., 160.

[13]Mary Burns, "RDA and Rare Books Cataloging, Part 2," *Library Resources and Technical Services* 63, no. 1 (2019): 25, https://doi.org/10.5860/lrts.63n1.4

[14]*Descriptive Cataloging of Rare Materials (Books)*.

[15]RDA Steering Committee, "About RDA," accessed 27 May 2022, www.rda-rsc.org/content/about-rda.

[16]*Descriptive Cataloging of Rare Materials (Books)*.

[17]RDA Steering Committee, "RDA Frequently Asked Questions," accessed May 27, 2022, http://rda-rsc.org/content/rda_faq#8.

[18]Robert Maxwell and John Attig, "Reconsidering DCRM in the Light of RDA: A Discussion Paper," accessed 27 May 2022, RDA Steering Committee, "RDA Frequently Asked Questions," accessed May 27, 2022, http://rda-rsc.org/content/rda_faq#8.

[19]Ibid., 6-8.

[20]Todd Fell and Francis Lapka, "ISBD and DCRM into RDA: An Opportunity for Convergence?" *Cataloging and Classification Quarterly* 54, nos. 5-6 (2016): 282-290, https://doi.org/10.1080/01639374.2016.1190436.

[21]Ibid., 286-289.

[22]Ibid., 290.

[23]"Final report of DCRM-RDA Task Force," October 10, 2012, accessed 27 May 2022, 3, https://rbms.info/files/dcrm/rda/FinalDCRMRDAReport_20121010.pdf.

[24]Meeting minutes of RBMS Bibliographic Standards Committee at ALA Midwinter Conference, 26 January 2013, accessed 27 May 2022, 7-8, https://rbms.info/files/committees/minutes/2013/bibstandminutes13m.pdf.

[25]*BIBCO Standard Record (BSR) RDA Metadata Application Profile* (Washington, D.C.: Program for Cooperative Cataloging, 2020), accessed 27 May 2022, 4, www.loc.gov/aba/pcc/bibco/documents/PCC-RDA-BSR.pdf.

[26]Meeting minutes of RBMS Bibliographic Standards Committee, 26 January 2013, 12.

[27]Francis Lapka, email message to DCRM Users' Group mailing list, 18 April 2013, https://listserver.lib.byu.edu/pipermail/dcrm-l/2013-April/003138.html.

[28]Meeting minutes of RBMS Bibliographic Standards Committee at ALA Annual Conference, 29 June 2013, accessed 27 May 2022, 5-6, https://rbms.info/files/committees/minutes/2013/bibstandminutes13a.pdf; Meeting minutes of RBMS Bibliographic Standards Committee at ALA Midwinter Conference, 25 January 2014, accessed 27 May 2022, 5, https://rbms.info/files/committees/minutes/2014/bibstandminutes14m.pdf.

[29]"ACRL/RBMS Descriptive Cataloging of Rare Materials Task Force," Rare Books and Manuscripts Section, Association of College and Research Libraries, accessed 27 May 2022, https://rbms.info/acrl-rbms-descriptive-cataloging-of-rare-materials-task-force/.

[30]Ibid.

[31]Ibid.

[32]Meeting minutes of RBMS Bibliographic Standards Committee at ALA Annual Conference, 24 June 2017, accessed 27 May 2022, 6, https://rbms.info/files/committees/minutes/2017/bibstandminutes17a.pdf; Meeting minutes of RBMS Bibliographic Standards Committee at ALA Midwinter Conference, 10 February 2018, accessed 27 May 2022, 7-8, https://rbms.info/files/committees/minutes/2018/bibstandminutes18m.pdf

[33]Robin Camille Davis, "Git and GitHub for Librarians," *Behavioral & Social Sciences Librarian* 34, no. 3 (2015): 163-164, https://doi.org/10.1080/01639269.2015.1062586.

[34]Ibid., 158.

[35]Prayudi Utomo and Falahah, "Building Serverless Website on GitHub Pages," *IOP Conference Series: Materials Science and Engineering* 879 (2020): 4, https://doi.org/10.1088/1757-899X/879/1/012077.

[36]Ibid., 1.

[37]Yasset Perez-Riverol, et al., "Ten Simple Rules for Taking Advantage of Git and GitHub," *PLoS Computational Biology* 12, no. 7 (2016): 1, https://doi.org/10.1371/journal.pcbi.1004947.

[38]Keith Engwall and Mitchell Roe, "Git and GitLab in Library Website Change Management Workflows," *Code4lib Journal* 48 (2020), https://journal.code4lib.org/articles/15250.

[39]Ibid.

[40]Meeting minutes of RBMS Bibliographic Standards Committee at ALA Midwinter Conference, 26 January 2019, accessed 27 May 2022, 10, https://rbms.info/files/committees/minutes/2019/bibstandminutes19m.pdf.

[41]Report of RBMS RDA Editorial Group, 27 April 2020, accessed 27 May 2022, 1, https://rbms.info/files/committees/bibliographic_standards/committee-docs/RBMS%20RDA%20Editorial%20Group%20Report-200423.pdf.

[42]Ibid.

[43]Creative Commons, Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0), accessed 26 May 2022, https://creativecommons.org/licenses/by-nc-sa/4.0/.

[44]Google, "What is Programmable Search Engine?" accessed 9 May 2022, https://developers.google.com/custom-search/.

[45]RBMS RDA Editorial Group, "Citation numbers," revision 8, last modified 1 December 2021, accessed 27 May 2022, https://github.com/rbms-bsc/DCRMR/wiki/Citation-numbers.

[46]RBMS RDA Editorial Group, "Style sheet (full)," revision 28, last modified 7 March 2022, accessed 27 May 2022, https://github.com/rbms-bsc/DCRMR/wiki/Style-sheet-(full).

[47]Ibid.

[48]Ibid.

[49]GitHub, "The largest open source community in the world," accessed 29 April 2022, https://github.com/open-source.

[50]Liz Adams and Francis Lapka, "Hosting Options for DCRM2," report presented at the 2019 ALA Midwinter Conference, accessed 29 April 2022, https://docs.google.com/document/d/1J150DeeDxdFSK_ohC8pfCHWy32D_-iImgFt8cw1-e4g/edit.

[51]Society of American Archivists' Technical Subcommittee on Describing Archives: A Content Standard (TS-DACS), "Describing Archives: A Content Standard," Version 2021.0.0.2, last modified 28 May 2021, accessed 27 May 2022, https://github.com/saa-ts-dacs/dacs/blob/master/dacs.md.

[52]Git for Windows, revision 2.36.0, accessed 29 April 2022, https://gitforwindows.org/.

[53]Ruby, revision 3.1.2, accessed 29 April 2022, www.ruby-lang.org/en/.

[54]Jekyll, revision 4.2.2, accessed 29 April 2022, https://github.com/jekyll/jekyll.

[55]Minimal Mistakes, revision 4.24.0, accessed 27 May 2022, https://github.com/mmistakes/minimal-mistakes.

[56]RBMS RDA Editorial Group. "Updating the Jekyll theme & customizations," last modified 26 January 2022, accessed 27 May 2022, https://github.com/rbms-bsc/DCRMR/wiki/Updating-the-Jekyll-theme-&-customizations.

[57]Don Ho, "Notepad++," revision 8.4, accessed 27 May 2022, https://notepad-plus-plus.org/.

[58]Anaconda, "Who is Anaconda?" revision 3.9, accessed 29 April 2022, www.anaconda.com/about-us.

[59]RBMS RDA Editorial Group, “rbms-bsc/DCRMR GitHub Repository,” last modified 16 February 2022, accessed August 24, 2022, https://github.com/rbms-bsc/DCRMR/.

[60]RBMS RDA Editorial Group, "Welcome to the DCRMR Editorial Guidelines Wiki," last modified 14 March 2022, accessed 27 May 2022, https://github.com/rbms-bsc/DCRMR/wiki/.

[61]DCRMR implemented Google Analytics 4 (GA4); pre- and post-GA4 statistics are maintained separately.

[62]RBMS RDA Editorial Group, "Issues," last modified 12 October 2021, accessed 27 May 2022, https://github.com/rbms-bsc/DCRMR/issues.

[63]RDA Steering Committee, "Frequently Asked Questions."

[64]RBMS RDA Editorial Group, “rbms-bsc/DCRMR/CITATION.cff,” last modified 20 May 2023, accessed 31 January 2024, https://github.com/rbms-bsc/DCRMR/blob/main/CITATION.cff.

[65]Descriptive Cataloging of Rare Materials (RDA Edition) (v2022.2.1.1), https://doi.org/10.5281/zenodo.7953732.

[66]Descriptive Cataloging of Rare Materials (RDA Edition) (v2022.2.0.1), https://doi.org/10.5281/zenodo.7921077.

Two years at the British Library…and counting.

Alan Danskin
Former Collection Metadata Standards Manager, British Library

Introduction

My first CIG conference was at Retford, about 30 years ago. I am very honoured to have been invited to give the keynote today.

"[Collection metadata] is a key organisational asset, representing centuries of resource investment"

Figure 1. British Library Collection Metadata Strategy documents

If this presentation has a theme, it is that collection metadata is a key organisational asset for any library. This has been the core message of the British Library's Collection Metadata Strategies from 2015 to the present day. Obviously, metadata is essential for discovery and management of collections and, as was recently acknowledged by George Osborne, uncatalogued collections are not secure.

The title of the presentation refers to my intention that, after a couple of years working as a cataloguer at the British Library, I would leave and return to Scotland to work in a university library. Thirty-six years later I have retired from the British Library. So, what went wrong?

When I joined the British Library in 1987. The British Library itself was still a relatively new institution. It was just 14 years old. The St Pancras site was still shrouded in hoardings and the collections were distributed over at least 10 sites spread across London and at Boston Spa in West Yorkshire.

The Library then was very different from the library today. The management style was very hierarchical. Information was communicated very much on a need-to-know basis. It was generally exclusive. The reference services located at the British Museum were regarded as a library of last resort. It was difficult to get a reader's ticket. Some services, such as Science Reference Information Service (SRIS), were open access. There were conflicts between the free to access public services and the commercial imperatives of the British National Bibliography and Document Supply Centre. The physical separation of staff and collections across the Library's estate meant that old loyalties died hard. My manager, Pat Oddy (second from the left in the photograph below), was inclined to compare the different directorates to mediaeval fiefs ruled by competing robber barons.

The world of cataloguing was also very different. It surprised me that cataloguers didn't do direct data entry. Pat Oddy had a desktop computer for word processing, but cataloguers had fiche reader for routine

consultation of BNB and Precis subject packages. Catalogue records were written by hand on UKMARC input sheets and sent to a keying bureau in Birmingham, before being dispatched on tape to the IBM mainframe in Harlow that the Library rented from Rank Hovis McDougall. There were no mobile phones, no World Wide Web, but we could look up the catalogues for post 1975 collections and the British National Bibliography on BLAISE (the British Library Automated Information Service) – which you can see displayed on the monitor in the photograph below.

Figure 2. The British Library Catalogue Conversion Project team (1990). The Project Manager, Pat Oddy is second from left. ©

I was assigned to the project to convert the British Library Catalogue of Printed Books to UKMARC. This photo was taken in the Novello Room at Novello House on Sheraton Street in Soho, which was where the Bibliographic Services Division was based. As you can imagine, Soho was an interesting and lively place in which to start working in London – lunchtime drinking was not compulsory, but it was considered a normal thing to do.

The past may be another country and we did drink differently there. But, as William Faulkner put it, “The past isn’t dead. It isn’t even past.” While much has changed since 1987, the legacy of the past lives on and a large part of my career has been spent trying to address the deficiencies of the metadata the Library inherited with its disparate collections and services. So, I want to start where I began, with the project to convert the British Museum catalogue of printed books into machine readable format – MARC.

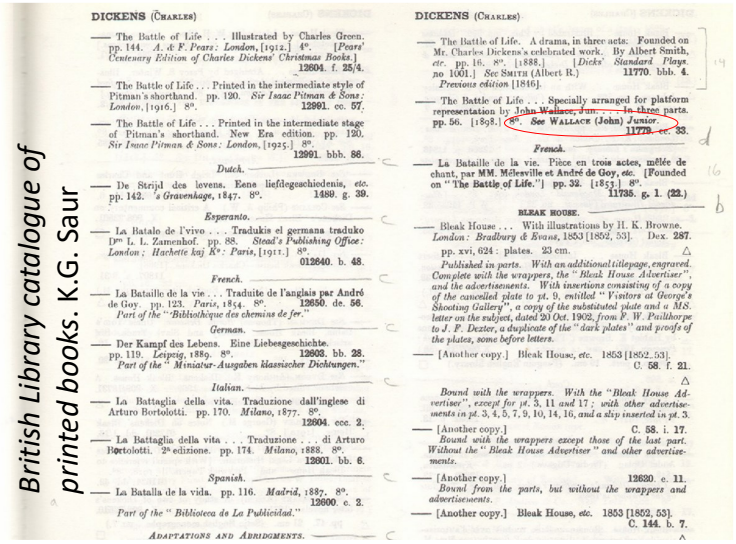


Figure 3. Detail of a page from British Library Catalogue of Printed Books. This is from the conversion team's copy and contains our mark-up and annotations

British Library Catalogue Conversion Project

In the late 1980s the Library was planning for the move to St Pancras. However, there was no room at St Pancras for GK (the General Catalogue) whose guard books took pride of place in the centre of the round reading room at Bloomsbury. The decision had been taken to convert the catalogue into UKMARC format so that it could be consulted through computerised catalogue.

The source for the work was the 360 volume *British Library Catalogue of Printed Books* and its six supplements, published by K.G. Saur. This catalogue was the final printed manifestation of the catalogue of the British Museum Library.

The catalogue makes wonderfully effective use of the print technology of its time to render the complexity of publications intelligible to its users. Entries are arranged by alphabetical sequence of authors and titles. Different editions of the same work are gathered and collocated with translations and other related works. But the structure is implicit in the arrangement on the page. The structure was exploited to render the catalogue into UK MARC and later MARC 21.

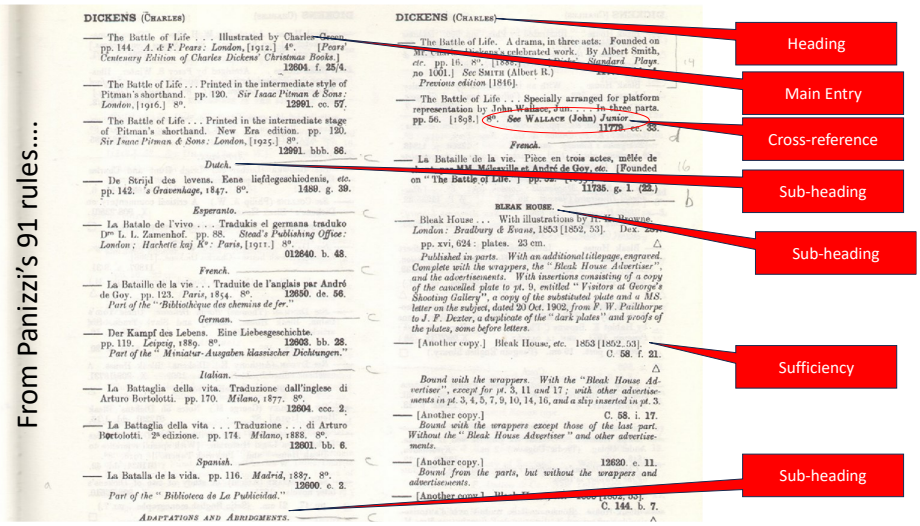


Figure 4. Illustrating the structure of BLC.

The catalogue rules used here are those of the British Museum, developed from Panizzi's 91 rules. By adding some labels, we can see the structure more clearly. At the top of each column is the heading. Followed by a series of main entries for publications in English of the *Battle of life* by Charles Dickens. Subheadings for other languages follow. Then another subheading for adaptations and abridgements, which we would think of as added entries pointing to the main entry under the name of the person who adapted it. Then another subheading introduces a different work, *Bleak house.*

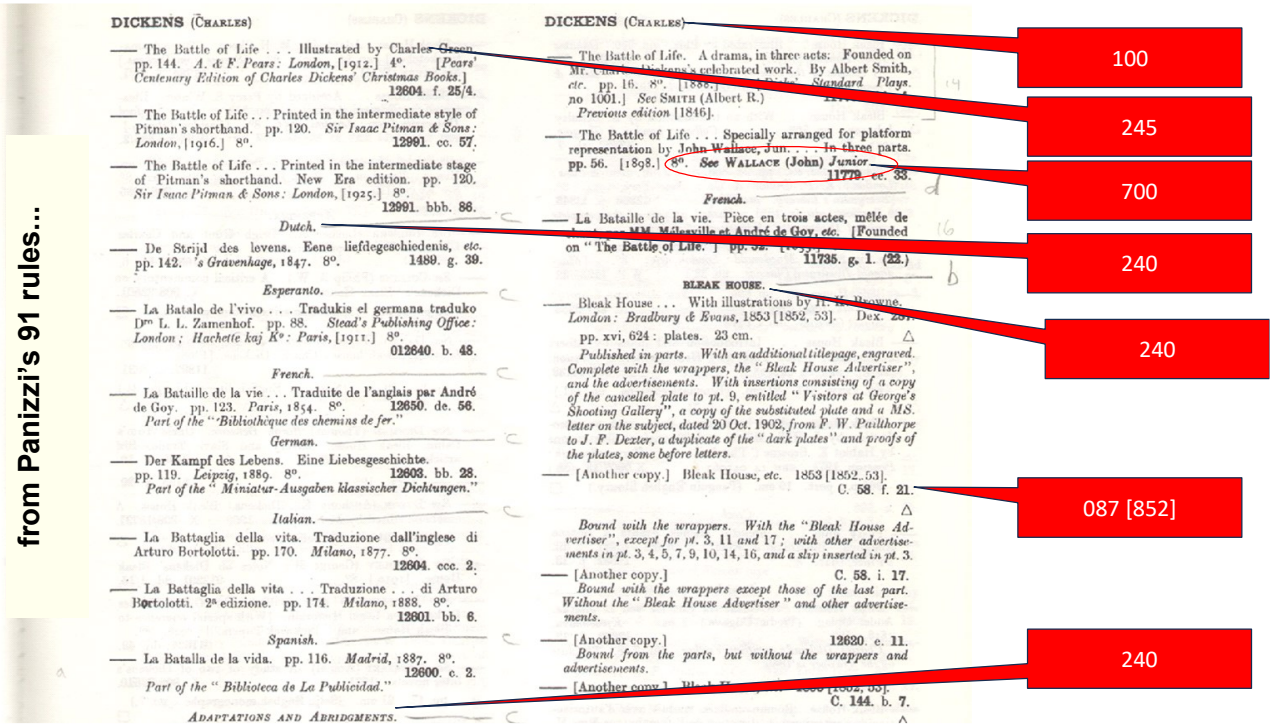


Figure 5. Illustrating assignment of MARC encoding.

BLC MARC (a simplified form of UK MARC) was used to encode the data so that the BLC file could be made available on BLAISE in the BL OPAC. I will return to this later, but in 2004 the BLC was one of about twenty UKMARC files migrated to the Library's new Library Management System, Aleph. As part of that process, the BLC MARC data had to be converted to MARC 21, as Aleph did not support UKMARC.

The conversion to MARC was obviously an essential step, but the flat structure of MARC traded the collocation and sufficiency of the printed catalogue for efficient data exchange and keyword searching. By the mid to late 1990s the impacts of digitisation and the World Wide Web had begun to highlight limitations of both MARC and the Anglo-American Cataloguing Rules, leading to the development of *Functional Requirements for Bibliographic Records* (FRBR), *RDA: resource description and access*, and eventually the *IFLA Library Reference Model* (LRM).

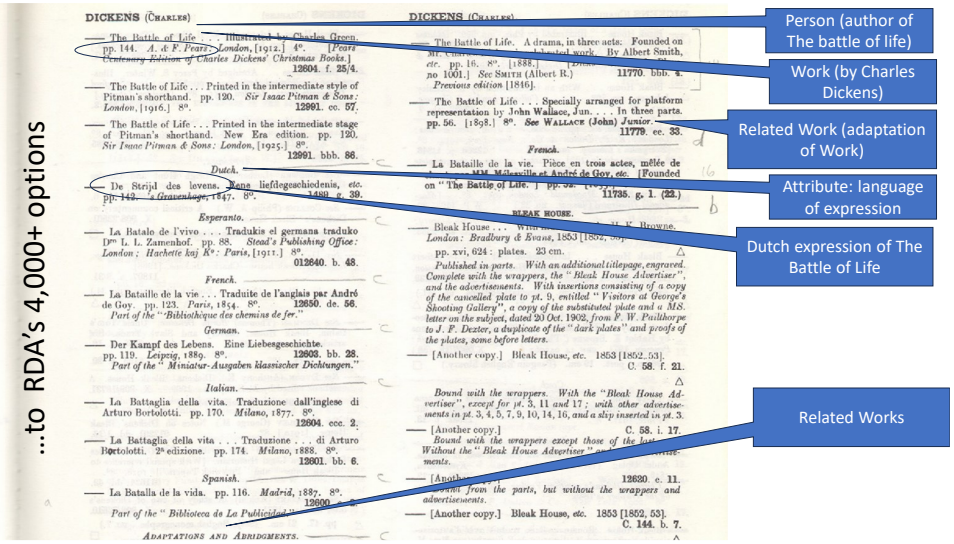


Figure 6. Illustrating alignment with RDA.

If we review the BLC example above, it is easy to see how its structure can also be expressed in RDA. Charles Dickens is a Person; *The battle of life* is a Work. "Charles Dickens is the author of *The battle of life"* expresses a relationship between a Person and a Work. The subtitles help to identify multiple language expressions of *The battle of life* and each description includes attributes of the Manifestation and Item.

Other subheadings also provide evidence for relationships between the original work and derivative works, such as adaptations. FRBR, LRM and RDA, represent a return to a more user centred approach by making a clear separation of content from carrier and by the relative weight given to relationships placing the resource described in context within the collection and the wider universe of human discourse.

Enhancement of Legacy Metadata

Figure 7. The "Title room" British Museum pre-1914.

These gentlemen were the custodians of the Library's metadata in the Edwardian era. On the shelves around them are millions of title slips, which are the British Museum's original catalogue records. The title slips have since been moved to new accommodation at Boston Spa and are still consulted regularly to answer queries from staff and the public.

"I. Titles to be written on slips, uniform in size."
Rules for the compilation of the catalogue

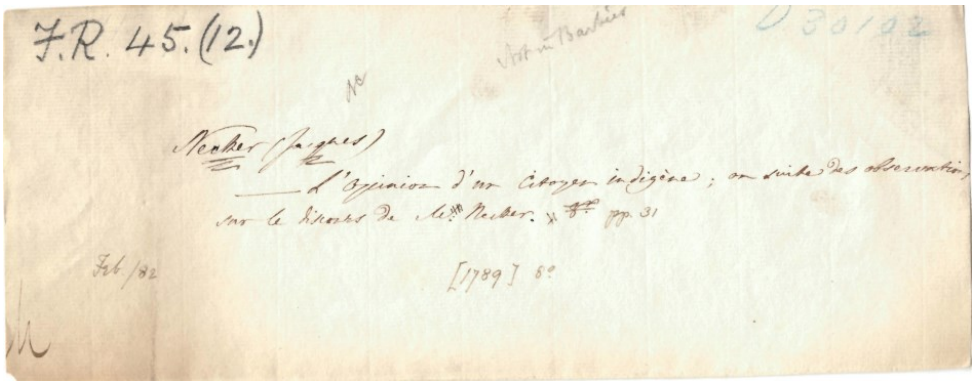


Figure 9. Title slip, attributed to Sir Anthony Panizzi

FMT	BK
LDR	nam a22001573 4500
001	002614806
005	20201009183539.0
008	890802|1789|||||| || ||||fre
040	|a Uk |c Uk
1001	|a Necker, Jacques, |d 1732-1804.
24010	|a Appendix
24510	|a L'Opinion d'un citoyen indigène; ou suite des observations, sur le discours de M. Necker.
264 1	|c [1789]
300	|a 31 pages ; |c (8º)
336	|a text |2 rdacontent
337	|a unmediated |2 rdamedia
338	|a volume |2 rdacarrier
340	|m 8vo |2 rdabf
85241	|a British Library |b HMNTS |j F.R.45.(12.)
SYS	002614806

Figure 9. Current MARC record

This is the current MARC 21 record, copied from Explore the British Library. As you can see, there is quite a lot more data here than on the original slip. So, what has changed? When did it change? And why has it changed?

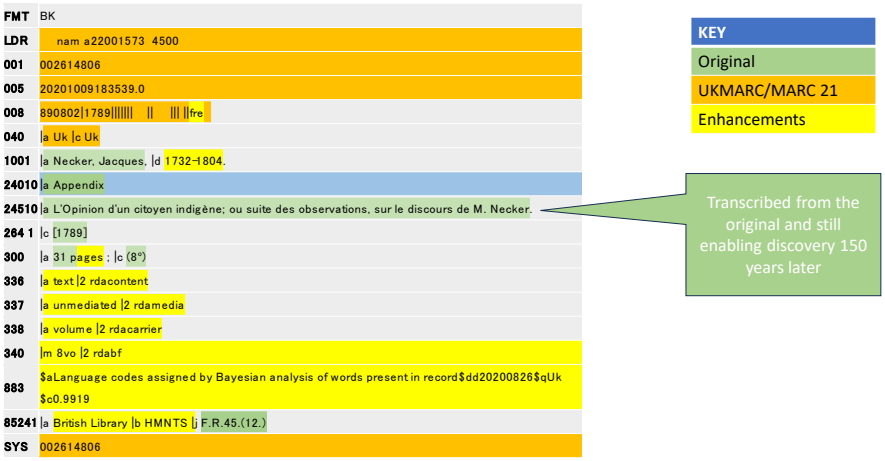


Figure 10. Illustrating legacy transcribed data in green.

The original data is still there, highlighted in green. Panizzi's instruction to transcribe from the book in, "as few words [as necessary] and those only of the author" meant that the data was accurate and remains effective more than a century after it was created. That seems like a decent return on investment.

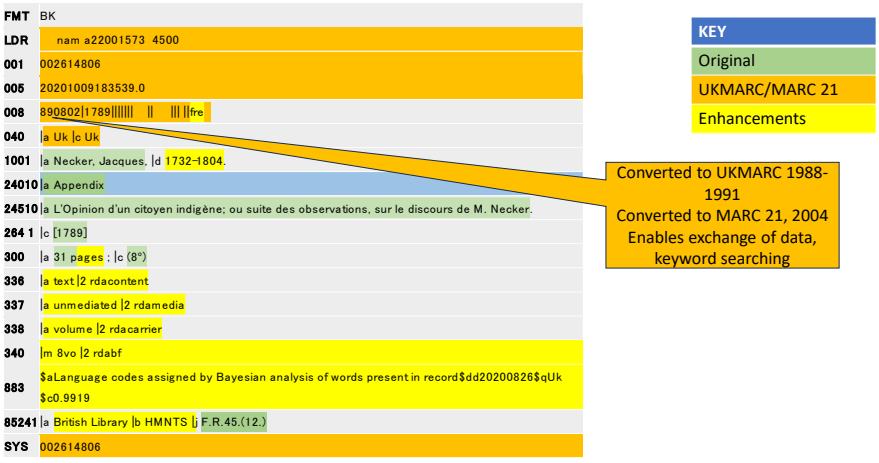


Figure 11. Illustrating conversion to MARC.

In converting the record to UK MARC and later to MARC 21, the encoding was added, as well as various data elements required for data exchange, such as system number, leader, latest transaction, and source of metadata. BLC MARC proved quite difficult to transform to MARC 21 because of the ambiguity of some of the simplified encoding. For example, all main entries were encoded 100 and the whole publication statement was bundled into 260 $a (anticipating RDA's manifestation statement). The simplification was probably an essential component in the success of the BLC Conversion project, but it is another example of how decisions taken for perfectly good reasons can have consequences down the line.

Alignment with North American Standards was a major component of the library's cataloguing strategy, devised by to improve efficiency Pat Oddy in the early 1990s. Replacing local authority files with NACO and

LCSH enabled reuse of Library of Congress and OCLC records with minimal intervention. MARC 21 was also essential for compatibility with new library management systems, which did not have native support for UKMARC.

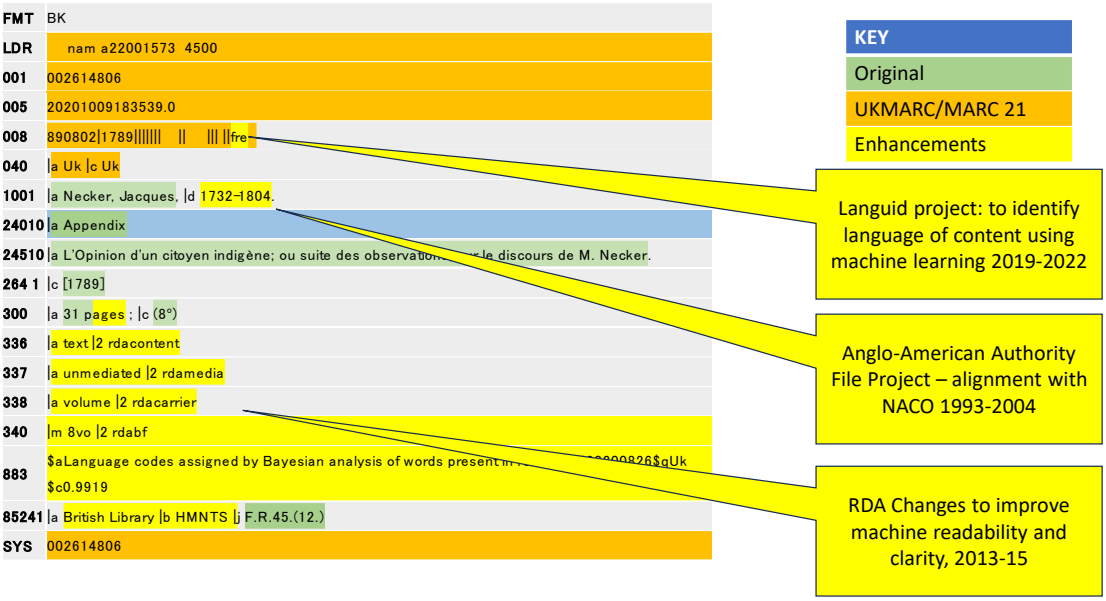


Figure 12. Illustrating enhancements over time.

Beginning in 1993, the Anglo-American Authority File was a collaboration with Library of Congress by which the Library was able to join NACO. Following the implementation of Aleph, Richard Moore ran several projects to align internal name headings with NACO. This greatly improved collocation of legacy records with more recent records created under AACR2. The focus for these projects was to convert the most common headings, so emphasis was on best-selling or canonical authors, sacred works, and corporate bodies. Richard gave his team some leeway regarding choice of headings that they worked on to maintain enthusiasm and quality. During this period, Andrew MacEwan led the implementation of LCSH to replace the unlamented CompASS subject vocabulary.

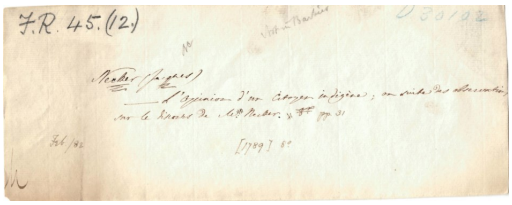
More recently, Thurstan Young undertook a substantial project in 2014/15 to assign RDA content, media, and carrier types to legacy records. Other RDA influenced changes can be seen in the expansion of abbreviations in the 300 field to make the catalogue more accessible to non-cataloguers, we also took advantage of the expansion of MARC 21 to copy book format from the extent to a specific subfield in 340.

Another major enhancement has been the assignment of language codes in 008/35-37 using statistical methodologies and machine learning, which I want to describe in a bit more detail. As well as the language code, note that the 883 field contains provenance information for the project, including the confidence level for each assignment recorded in $c.

Languid Project 2018-2022

This book is in French

- Implicit to humans
- Opaque to computers
- Estimated 4.5 million records lacked MARC language codes
- Mostly pre-1975 collections



Title: L'Opinion d'un citoyen indigène; ou suite des observations, sur le discours de M. Necker.
Author: Jacques Necker, 1732-1804.
Publication Details: [1789]
Language: French
Uniform Title: Appendix
Identifier: System number: 002614806
Physical Description: 31 pages ; (8°)
Shelfmark(s): General Reference Collection F.R.45.

Figure 13. Languid project

The Languid project addressed the longstanding problem of identifying the language of items in the collection. With the exceptions of canonical authors, official publications and sacred works, the language of the expression was not recorded in legacy records because it was implicit to users and there was no way to search a linear name-title catalogue by language. I estimated from our annual audit of the catalogue metadata that around 4.5 million records lacked this information.

The project began at the end of 2018 and was the work of Victoria Morris – who will be speaking later today on another project. Victoria used a machine learning approach to identify the probability of words in the title belonging to a specific language. This was possible because we already had millions of records which contained accurate language codes. Using this corpus, the machine could be trained to recognise languages and assign a confidence level. Setting a high confidence level was essential because, with such a large dataset, assignment errors would be very hard to detect. Victoria was also able to engage the aid of Library's community of language experts to resolve ambiguous results. This proved to be a very popular activity during lock down, when curatorial staff and cataloguers were unable to access the collections and it meant that the confidence level for assignments was well above 99%.

Languid Project 2018-2022

Dependencies

- Transcription of title
- Corpus of accurate metadata
- Expert professionals
- Capacity to try things out

Benefits:

- Improved knowledge of collections
- Better discoverability
- Inclusivity

"XVIII The title of the book next to be written, and that expressed in as few words, and those only of the author, as may be necessary to exhibit to the reader all that the author meant to convey in the titular description of his work; the original orthography to be preserved." A Panizzi
Rules of the compilation of the catalogue. 1841

Figure 14. Languid project overview

This approach depended absolutely on the Anglo-American cataloguing tradition of accurate transcription from the title page, which can be traced back to Panizzi's codification of cataloguing rules in 1841. The availability of good quality metadata from which to create a language corpus was also vital. Nothing could have been achieved without the enthusiasm, technical competence, and linguistic expertise of the staff involved.

The benefits are substantial. Curators covering the African and Australian collections have commented that it has enabled them to get a much clearer picture of strengths and weaknesses of the collections for which they are responsible. Discoverability is improved, searches refined by language will achieve a much better recall and precision. Better understanding of the language content of the collection also enables us to serve different language communities more effectively. I hope that the more granular identification of minority and endangered languages will allow us to reach out to audiences who may have felt excluded in the past.

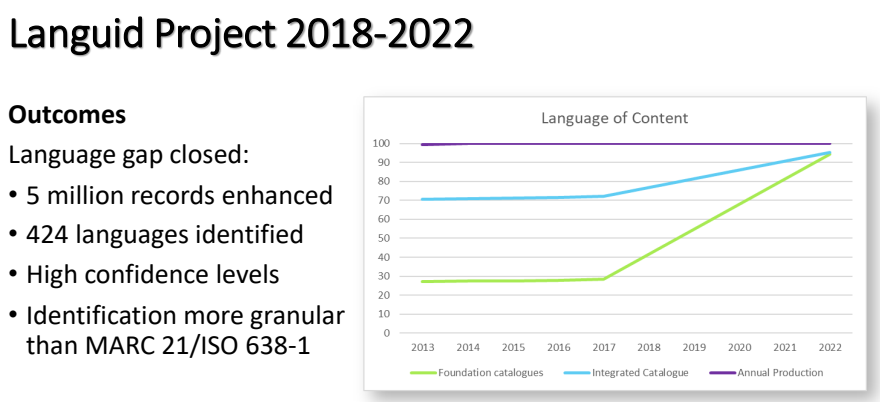


Figure 15. Closing the language gap.

This graph from the annual audit of the catalogue illustrates the quantitative outcome from the project. The blue line along the top shows that 100% of new MARC records produced each year contain a valid language code. The green line at the bottom represents "foundation collections", including the British Museum catalogues, only about a quarter of which contained language codes prior to the Languid project. As shown by the pale blue line in the middle the gap between language coverage for current material and the catalogue as a whole was 30% in 2013, when we started monitoring this. The gap between current practice the foundation collections was more than 70%. As you can see, by end of 2022 the gap had closed to less than 10%.

Languid resulted in the assignment of million language codes to more than 5 million records, representing all of the 400 + languages identified by the MARC 21 language codes. We hoped that the methodology would be applicable to other catalogues and collections, including our sound recordings and our archives and manuscripts. Unfortunately, in those catalogues title information is much less likely to have been transcribed.

There is always more that could be done. It would be nice to make use of the more granular identification of languages by assigning codes from ISO 639-3 to supplement the MARC 21 codes for language groups and I hope it will be possible to take this forward through future work on equity and inclusion.

We have made a modest start towards addressing outdated or inappropriate demonyms by updating LCSH headings containing the term blacks. There is much more to do, and work is under way on evaluation of terminological and provenance issues in the Caribbean and South Asian collections.

The introduction of FAST to replace LCSH was motivated in part by the need for a more agile approach to subject indexing. Subject coverage across the catalogue is far from comprehensive, with the legacy data being particularly weak. This is unfortunately not easily addressed by artificial intelligence. During lockdown we were able to map FAST headings to the Watts *Elastic Classification* used by the British Museum and assign subject and genre terms to thousands of records. There are obviously limitations to how far contemporary terms can be mapped to 19th century classifications, but we discovered that by far the biggest problem, was that what was actually shelved at specific locations did not always correspond with the classification. We also benefit from the enrichment of our 19th century books data base by UCL students on the *Information through the ages* course, who have contributed thousands of subject, genre and geographical terms since 2019. I hope that much more can be done by matching and merging our legacy records with the richer records in the *English Short Title Catalogue*.

Further projects combining the *Georeferencer* tool with the wisdom of a select crowd of enthusiasts have enabled us to add coordinates and geographic headings to cartographic materials to improve visualisation of the collection in future.

In 2022, we took the decision to supersede LCSH with FAST as our preferred subject vocabulary. This was not an easy decision. One of the benefits of FAST is that each term has a unique identifier, which can be expressed as a URI (Uniform Resource Identifier). Encoding information as URI's makes it machine actionable and provides new ways of linking and reusing metadata. In addition to FAST, we have also begun to embed ISNI's in our bibliographic records, as a result of a successful collaboration between the Library, major publishers and our Cataloguing in Publication agent, BDS.

Linked Data and the Share Family

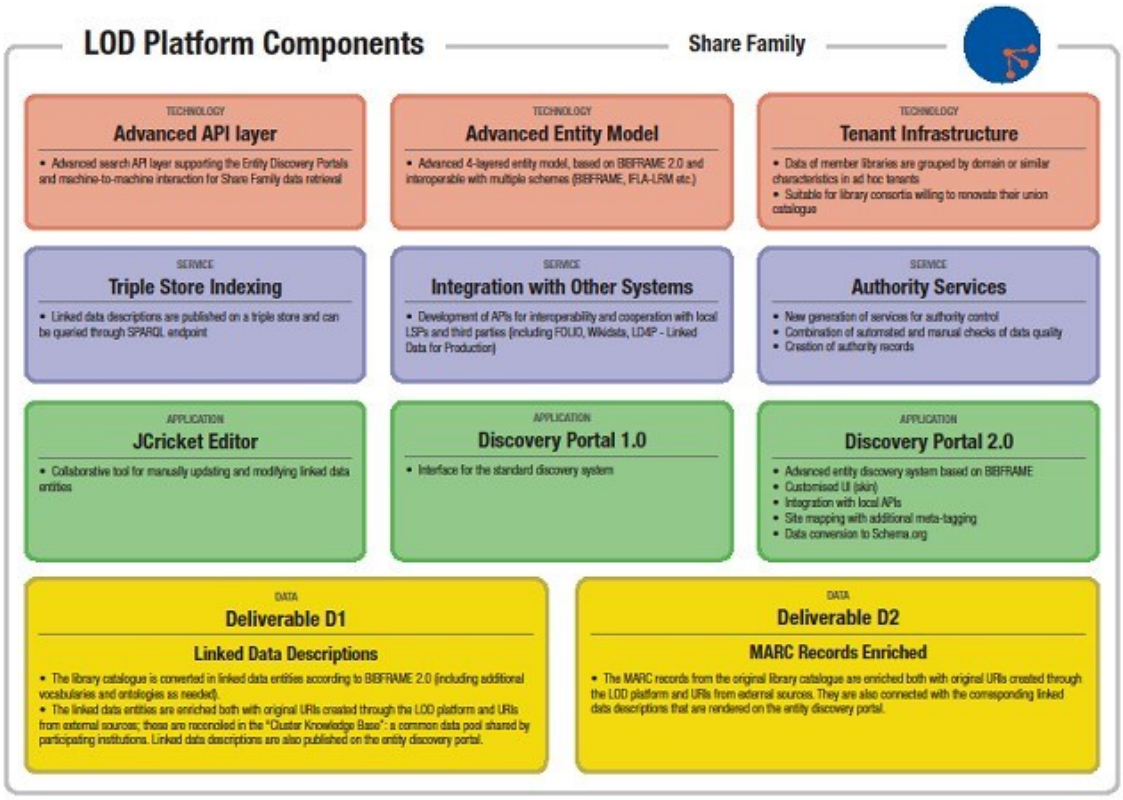


Figure 16: Overview of Share-VDE

The Library began to explore linked data in 2010. The British National Bibliography (BNB) was published as

linked open data the following year in collaboration with Talis and later TSO. However, the Library lacked the capacity to develop the service and in 2020 we took the decision to join the Share-Family. The Share Family is a collaborative community led by Casalini Libri and @Cult to develop a shared virtual discovery platform for Libraries. The "parents" are based in Italy, but it is a global family. Many of the members have been involved in Bibframe development through LD4P, etc. The Library was concerned that Share-VDE's data model should support LRM/RDA data as well as Bibframe and has contributed to the development of the Share data model. This is still in development, but a major step was the introduction of svde:Opus to enable reconciliation of bf:Work with rda:Work and rda: Expression.

Figure 17: BNB home page on Share platform

BNB data has now been loaded to Share-VDE and BNB is the first member of the national bibliographies' tenancy. The interface went live in June, shortly before I retired. The BNB has its own search page, but it is a part of the underlying knowledge base. The Share Family provides the potential to aggregate and cross search multiple national bibliographies for the first time.

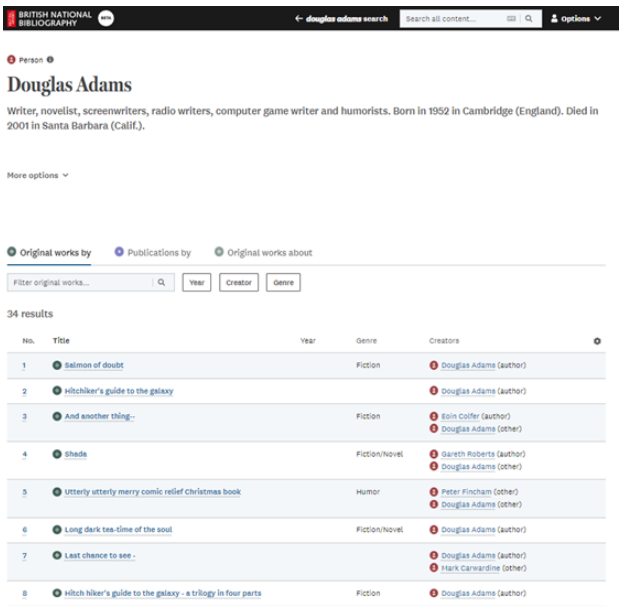


Figure 18: Example search results for original works by Douglas Adams

The Library expects to benefit from Share-VDE by learning from the practical experience our collaborators have gained through other linked data initiatives and by gaining access to a state of the art, extensible linked data infrastructure designed for library data. In Share VDE we will have a new channel for dissemination of the BNB, in aggregation with other national bibliographies.

Share-VDE also provides an opportunity to re-tool our metadata for the 21st Century. Our bibliographic metadata will be remodelled and clustered making it more compatible with entity-based data models, including the IFLA Library Reference Model, RDA: Resource Description and Access, and Bibframe. Our data will be enriched with URIs which will make it more effective in linked data environments. An entity-centred view of the BNB and subsequently the British Library catalogue will offer new perspectives for researchers.

Quality Assurance Metadata Measure (QAMM)

I've talked a lot about good metadata, but what is good metadata? This is another issue that we have had to address over the years. Cataloguing is often seen as a luxury that can be cut back or done without when times are hard. Cataloguers are sometimes accused of gilding the lily or creating gold plated metadata. During lockdown I began to work with my Quality Assurance team to develop a new measure for metadata quality (QAMM), as part of a wider review of key performance indicators.

The following design principles were agreed.

- A method that is extensible across all collections and workflows: the method based on FRBR was too closely tied to MARC to be utilised by archives and MSS for example.
- A method that enables valid comparisons between collections and workflows: It's difficult to compare inputs from different workflows and standards, so we decide to compare outcomes.
- A method that enables quantification of the value added by the Library: What difference does our intervention make to discovery or collection management or efficiency.
- A method that accounts for stewardship of the Library's metadata assets: Enables us to identify the impact of different interventions over time.
- A method that identifies issues requiring remediation: Is sufficiently granular to enable us to identify and intervene where metadata falls short of the requirement.

We looked at a number of existing quality methodologies, but they tended to be MARC based and aimed at identification of candidate records for copy cataloguing, so we developed a different approach based on outputs, rather than inputs. An approach focused on what the description could deliver for the user rather than what the cataloguer had delivered in accordance with standards.

Fitness for purpose

"Purpose and uses of the catalogues" in section 6 of *IFLA International Cataloguing Principles* (ICP)

- to **FIND** bibliographic resources in a collection as the result of a search using attributes or relationships of the entities
  - Single item (known item)
  - Set of resources by shared criteria
- to **IDENTIFY** a bibliographic resource or agent (that is, to confirm that the described entity corresponds to the entity sought or to distinguish between two or more entities with similar characteristics)
- to **SELECT** a bibliographic resource that is appropriate to the user's needs (that is, to choose a resource that meets the user's requirements with respect to medium, content, carrier, etc., or to reject a resource as being inappropriate to the user's needs)
- to acquire or **OBTAIN** access to an item described (that is, to provide information that will enable the user to acquire an item through purchase, loan, etc., or to access an item electronically through an online connection to a remote source); or to access, acquire, or obtain authority data or bibliographic data;
- to navigate and **EXPLORE**
- within a catalogue, through the logical arrangement of bibliographic and authority data and the clear presentation of relationships among entities
- beyond the catalogue, to other catalogues and in non-library contexts

+

- to **MANAGE** the collection efficiently by supporting automation of processes and provision of management information

We established a number of use cases, based on IFLA's International Cataloguing Principles. It's not a coincidence that these are close to the FRBR and RDA user tasks. I added an extra task, called Manage to cover metadata that we need for internal purposes, including data exchange and management information.

Example of BASIC description use cases

Description must satisfy the use case

Encoding examples are illustrative rather than comprehensive

Additional guidance added in response to issues raised in testing

ID0	Use case	Condition	Encoding	Status	PASS/FAIL	Additional Guidance
B01	FIND by the title or name of the entity described		MARC 2455a	Mandatory		
B02	OBTAIN the item or access to the content using shelfmark or link		MARC 852;MARC 856	Mandatory		**Print:** Ensure that items for validation are obtained from collection storage or from Finishing output **Digital:** Validate on Explore the British Library to confirm that access to digital resources can be obtained. There may be several options for accessing the same publication.
B03	SELECT in accordance with restrictions on access or reuse		MARC 506;MARC 540	Mandatory if applicable		Rights metadata may be added in Aleph or in the Metadata Aggregator prior to publication to Explore the BL. It is therefore important to check EtBL for digital resources. Born digital and digitized items, including Google books, will have a rights statement (Access restrictions) e.g. "No access restrictions"; and Terms governing use: e.g."Terms of use for Google books".
B04	SELECT by content, media or carrier type		MARC 336,337,338	Mandatory		A weeky fix routine is applied to records loaded to Aleph. Currently LDR 17=5 records are excluded and selection is limited to LDR 18=a. The routine will not assign any values if it is unable to assign content type. Collection Metadata receives a report of records that are updated and those that are selected but fail update.
B05	IDENTIFY or SELECT by attributes appropriate to the type of resource	See specific tabs	As specified	Mandatory if applicable		Check basic criteria for the relevant type of resource -e.g. serial, cartographic, music

bl.uk 27

The table above gives a zoomed in view of some basic use cases. A basic record should be sufficient to find a known item and obtain access to the content. The MARC coding is indicative not definitive and different encodings can be provided for other schemas. The QA team tested the basic concept, and we added additional contextual guidance to clarify ambiguities based on their feedback.

The measure was implemented in 2021 to supersede the previous "FRBR QM". There are currently three applications for the measure.

Corporate Quality Assurance measure (2021-)
- MARC Cataloguing is sampled and evaluated by the QA team.
- The intention is to roll it out to other systems and workflows over the next two years.

Performance Management (2023-)
- Team leaders sample and evaluate cataloguers' work.

- Same methodology but the emphasis in on using QAMM as a tool for coaching.

Outsourcing (2022-)
- QA team applies the measure to sample data provided by prospective suppliers.

QAMM is used in conjunction with validation tools, such as MARC Report. We are also looking at the different levels as a means of evaluation of what level of intervention is necessary or justifiable for specific types of material. For example, it was possible to quickly process low use educational material at basic level and send it to shelf. A similar technique is being evaluated for travel guides (of which we still receive many through legal deposit). A basic record with the addition of a geographical heading should enable discovery.

Conclusion

Figure 19. British Library Boston Spa aerial shot © British Library. Some of the buildings shown have been demolished to enable construction of a new high-density store.

So, what has changed since 2023 as the Library celebrates its 50th anniversary? There is "One British Library" occupying just two sites, at St Pancras and Boston Spa. Much has been standardised. The collection is treated as a whole, in so far as possible, although there is still no single catalogue covering the whole. The Library has become a much more public facing institution. It holds major exhibitions, and its services are far more standardised and accessible. The Library aims to be inclusive and to be relevant to diverse audiences in the UK and abroad. I've visited many national libraries over the years, but I've never seen any that rivals St Pancras for public presence.

In 1987 cataloguing was a self-contained operation, even the cataloguing in publication programme (CIP) was carried out in house. Since then, the Library has collaborated widely on standards development and metadata creation. This has been essential to maintain (and increase) productivity in line with increasing legal deposit. The importance of metadata has obviously grown over that period, although the contributions of those who create it are not necessarily recognised or valued to the same extent.

These are perceptions that we have tried, with some success, to change through our metadata strategy. The key message of the Collection Metadata Strategy, first published in 2015, is that Metadata is an institutional asset. Metadata is as important to the institution as its collection, estate, and staff. The metadata represents

the labour of many decades and even centuries of work by many hands both within the Library and beyond. Creating accurate, timely metadata is an investment, which will repay itself over time. However, metadata does not mature over time. It needs to be refreshed and enhanced to stay useful by responding to changing user expectations and exploiting new technologies.

Pat Oddy has a profound influence on my career and on my understanding or the public value of what we do. It is appropriate to end with a quote from her book *Future libraries future catalogues*, in which she sums up the value professional cataloguers and the benefit they confer on the Library user:

"Libraries provide access to the content of a collection in an organized and systematic way…The items in the collection are not only the organized output of human endeavour: the library also records and arranges these items in such a way as to give a further structural and organizational layer, which adds value to the whole and makes it far greater than the sum of its parts. This is a very long way from the provision of mere information and is at the heart of the professional values of the cataloguer."

Pat Oddy Future Libraries future catalogues p.11
London: Library Association Publishing, 1996.

The substance of this message remains as true today, as it was 27 years ago, or in Panizzi's day, but the potential of modern technology means that the scope of our activity need no longer be limited to individual collections or a single library, but can be a common legacy for all.

Further reading

British Library *Collection Metadata Strategy*
https://www.bl.uk/collection-metadata/strategy-and-standards

Alan Danskin (2020) "The Anglo-American Authority File: A PCC Story", *Cataloging & Classification Quarterly*, 58:3-4, 221-229, DOI: 10.1080/01639374.2019.1705952

Preprint: British Library Research Repository https://bl.iro.bl.uk/concern/articles/0c3236f9-47d1-4d9b-bce9-7e7b0303cdd6

Alan Danskin "*Challenging legacies at the British Library" Art Libraries Journal: Cataloguing Ethics Volume 48 Special Issue 2,* April 2023 , pp. 38 - 42 DOI: https://doi.org/10.1017/alj.2023.4

Preprint: British Library Research Repository https://bl.iro.bl.uk/concern/articles/66261da8-ccb6-4072-b8d1-e48b96f67d1d

Victoria Morris (2020) "Automated Language Identification of Bibliographic Resources", *Cataloging & Classification Quarterly*, 58:1, 1-27, DOI: 10.1080/01639374.2019.1700201

Preprint: British Library Research Repository https://bl.iro.bl.uk/concern/articles/6c99ffcb-0003-477d-8a58-64cf8c45ecf5

Janet Ashton & Caroline Kent (2017) "New Approaches to Subject Indexing at the British Library", *Cataloging & Classification Quarterly*, 55:7-8, 549-559, DOI: 10.1080/01639374.2017.1354345

Preprint: British Library Research Repository https://bl.iro.bl.uk/concern/articles/50a09642-7356-45de-b6c4-5724fecbef7c

Janet Ashton & Caroline Kent (2023) FAST: A Journey Toward Sustainability in Subject Indexing at the British Library, Cataloging & Classification Quarterly, 61:5-6, 525-534, DOI: 10.1080/01639374.2023.2238701

Pat Oddy Future Libraries future catalogues. London. Library Association. 1996 ISBN 1-85604-16-1

Sir Anthony Panizzi, *Rules for the compilation of the catalogue.* In: *Catalogue of printed books in the British Museum.* London. British Museum, 1841

(Also available online at Jeremy Norman's Historyofinformation website: https://www.historyofinformation.com/detail.php?entryid=2426 (viewed 1/9/2023))

Rules for compiling the catalogues of printed books, maps and music in the British Museum. London: British Museum, revised edition,1936.

"Purpose and uses of the catalogues" in section 6 of *IFLA statement of international cataloguing principles, 2016 https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2016-en.pdf*

UCL *Information through the ages https://www.ucl.ac.uk/module-catalogue/modules/information-through-the-ages-BASC0033*

Newbie cataloguer goes to the Metadata and Discovery Group Conference 2023. What happens next will shock you.

Jaycie Carter, Library Assistant, Birmingham City University

Last September I was very grateful to receive a sponsored place at CILIP's Metadata and Discovery Group (MDG) Conference 2023, taking place in my hometown Birmingham. I knew that the experience would be extremely informative – especially for someone who had only been cataloguing about a year and a half – but I didn't anticipate that I would also leave the conference feeling as inspired and welcomed to the profession as I did. The breadth and depth of topics discussed proved how diverse and rich a field metadata and discovery is. I learnt far too much to cover everything here, but what I will try to convey is how the conference helped give me a sense of the trajectory of cataloguing – both where we've come from and where we're heading next.

The conference actually started a day early for me, as I was involved in a tour of Birmingham for early arrivals – a wonderful way to meet a few friendly faces before the conference properly started! For the Birmingham City University (BCU) part of the tour, we showed delegates items from our University Archive and Special Collections Centre and chatted about the trials of cataloguing archival material in a Library Management System. I greatly appreciated hearing about how Dr Getaneh Alemu and Clare Delamore have innovatively dealt with this challenge in their respective institutions in their talks, giving us much food for thought at BCU.

Alan Danskin's keynote speech was an excellent way to kick off the conference, providing a concise overview of the history of cataloguing through the lens of his career at the British Library. Before this, I had not appreciated how much historical cataloguing techniques such as card cataloguing impacts our metadata today. For instance, it had never occurred to me that I see so many abbreviations in MARC records now because they were necessary for fitting as much information as possible on a handwritten card. Similarly, I was really interested to learn about the Languid project at the British Library. Cataloguing policy at the British Library ensured that titles were transcribed as written (i.e. rather than translated into English). Because of this decision, machine learning could be used to add language codes to 5 million records. As someone who has been apprehensive about artificial intelligence, this productive combination of up to 150 year old cataloguing policy and machine learning was heartening to see, and has made me feel much more excited about the possibilities machine learning can open up for our metadata and catalogues.

During his talk, Alan shared the words of Pat Oddy: "The items in the collection are not only the organized output of human endeavour: the library also records and arranges these items in such a way as to give a further structural and organizational layer, which adds value to the whole and makes it far greater than the sum of its parts" (1996: 11). One tool for creating this "structural and organizational layer" that I've thought about a lot before and since the conference are headings, which my table at the Cataloguing Code of Ethics workshop discussed in detail. The fourth statement of the code affirms that:

"interoperability and consistent application of standards help our users find and access materials. However, all standards are biased; we will approach them critically and advocate to make cataloguing more inclusive" (2022: 3).

Whilst the widespread use of Library of Congress Subject Headings (LCSH) makes it much easier to share records, we were all concerned about the well-documented biases within LCSH and how difficult and slow it can be to update offensive headings, especially when the Library of Congress has so much influence and power over decisions. There was a feeling that what we could achieve is limited when LCSH are so prevalent, essentially ensuring that information worldwide is being structured and organised according to the Library of Congress' worldview.

The issues with LCSH were further elucidated in two presentations. I was surprised to learn during Waseem Farooq's talk that native Bengali speakers often must learn the transliteration scheme used by the Library of Congress (e.g. for authority headings) in order to be able to access resources through libraries. Furthermore, Benjamin Cornish explored the challenges of trying to accurately describe items from the Natural History Museum's collection, in a way which accurately captures the colonial exploitation involved in natural history research expeditions. The cataloguer's decision about which information to include and exclude as well as which subject headings, authority headings, and relationship designators to use can all work to either reflect or obfuscate the history behind an item. The fact that the worldview, perspectives and contributions of indigenous people are not satisfactorily reflected in LCSH makes this task even harder. Hence, the Library of Congress has a concerning and deeply political influence over how we not only structure and organise, but also interpret and historicise the "organized output of human endeavour" (Oddy, 2022: 3).

That being said, I did not leave the conference feeling pessimistic by any means. The Cataloguing Code of Ethics highlights the importance of collaboration, which felt apt, as it was really energising talking to other cataloguers who care about ethical metadata in the workshop. Our table also discussed how helpful it is to build on other people's efforts. For instance, in their monthly blog *Critcatenate*, Violet Fox highlights any changes made to LCSH which are relevant to critical cataloguing; this list can help make it more manageable to update older, offensive LCSH without having to trawl through long PDFs, an approach we have considered doing at BCU. Martin Kelleher's work in setting up a UK NACO funnel – making it much easier for UK-based cataloguers to have an input in changes they want to see – further shows how committed the cataloguing community is to improving LCSH.

It was clear from the conference too that LCSH are by no means the only option available to cataloguers, and alternatives can at least be used to supplement LCSH, if not yet replace. In Jason Curtis and Lotty Summer's talk I learnt about how they have approached updating potentially offensive terminology in the Wessex Classification Scheme. Furthermore, Jenny Wright's informative talk really helped me understand the reasoning behind implementing FAST headings and how to go about using them. FAST headings were originally based on LCSH, but the process for petitioning changes to headings is supposed to be easier and faster. Since the conference, my colleague and I have begun using FAST headings at BCU as a result of Jenny's talk – as well as Homosaurus headings, which have now thankfully been added to Alma by Ex Libris!

As a newbie cataloguer, I'm sure it won't come as a surprise that I was apprehensive about the final part of the conference: the much-anticipated RDA Day. Whilst I had been made aware that I use RDA elements in my records and had seen the RDA Toolkit before, I had very little idea what was going on and was unsure whether the presentations would make much sense to me. I shouldn't have been worried, however. I was very grateful to the people on my table who answered my many questions (as well providing a really interesting discussion about linked data and the semantic web!). Furthermore, Gordon Dunsire, Thurstan Young, and Jenny Wright's presentations were all extremely useful, explaining the history of how RDA came about and its purpose.

I was interested to learn how RDA is a move away from AACR's emphasis on conveying the most important data in the most abbreviated form - originally to save space on handwritten cards, as mentioned in Alan Danskin's keynote speech. RDA offers cataloguers alternative choices which can improve discovery now that we're no longer so confined. On the other hand, RDA actually reduces the amount of information needed to meet the minimum requirements for a record. Libraries will likely continue to need a greater level of detail than this minimum requirement to meet their users' needs. However, this move towards economy opens up new possibilities for sharing metadata with organisations, groups and individuals outside of the library sector, who can use RDA to suit their own needs. Given the importance of collaboration in creating a more ethical metadata landscape, as affirmed in the Cataloguing Code of Ethics, I hope that RDA is able to open up new opportunities to share metadata more widely in the future.

Getting the chance to learn about all these new developments in metadata and cataloguing (as well as the hard work and dedication driving them behind the scenes) has made me feel like this is a really exciting time to be joining the profession. I'd really like to end this by thanking the MDG Committee for giving me this opportunity, but also all the delegates that I spoke to during the event. Everyone was happy to answer questions and explain acronyms and concepts to me. Moreover, I was always made to feel like my contributions to discussions were welcome and valued. It was certainly a little intimidating to begin with to be one of the least experienced people at the conference, but everyone was so welcoming and supportive that I was quickly put at ease (whilst perhaps not quite as shocking as the clickbait title would have you believe, it was a nice surprise for me!)

References

Oddy, P. (1996) *Future libraries future catalogues.* London: Library Association Publishing.

Cataloguing Ethics Steering Committee. (2022) *Cataloguing Code of Ethics*. Available at: https://docs.google.com/document/d/1IBz7nXQPfr3U1P6Xiar9cLAkzoNX_P9fq7eHvzfSlZ0/edit [Accessed 03 May 2024].

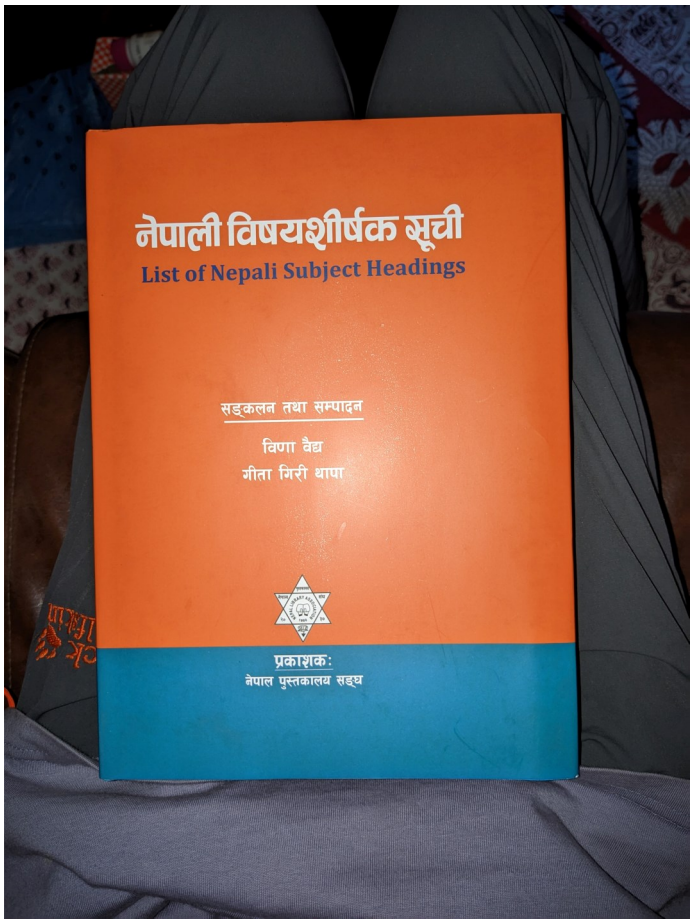
List of Nepali subject headings - an update

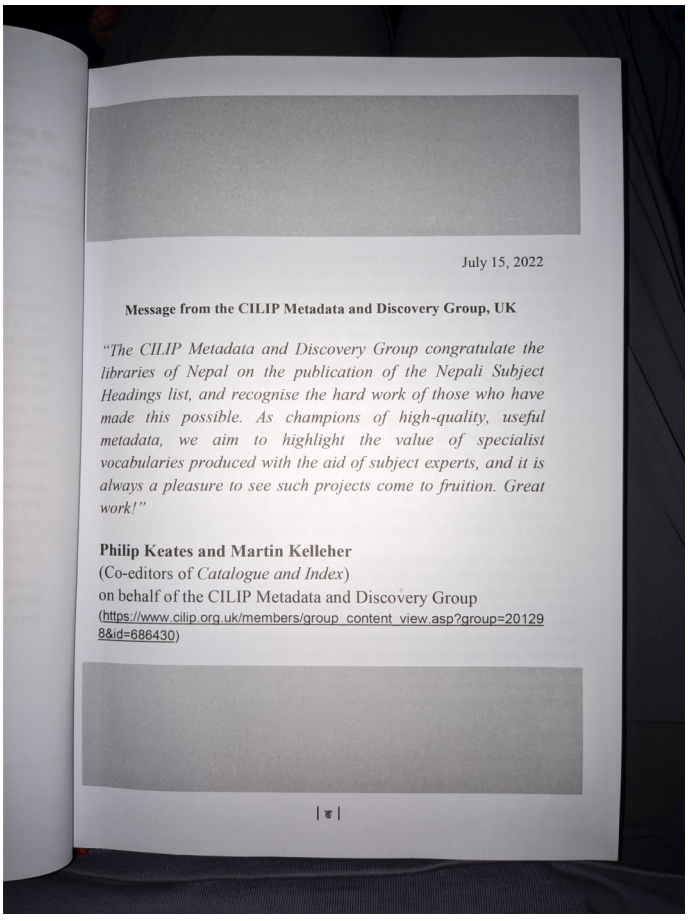
In issues 202 and 203 we published articles by Bina Vaidya and Gita Thapa reporting on their project to create and publish a list of Nepali subject headings.[1]

We recently received an update on the progress of the project:

“It is our pleasure to inform you that our book entitled "List of Nepali Subject Headings" has been published and launched by the Minister of Education, Science and Technology on Library Day in Nepal on September 1st, 2023. We have included five thousand three hundred Nepali Vocabularies with English translations and DDC Notation in one column. We have provided the introduction of the book in English to learn about the book by foreign readers. We will update the book soon. This is the first time the subject headings in Nepali language have been published from Nepal to bring uniformity in assigning Nepali subject headings

Bina Vaidya,
Gita Thapa”





July 15, 2022

Message from the CILIP Metadata and Discovery Group, UK

“The CILIP Metadata and Discovery Group congratulate the libraries of Nepal on the publication of the Nepali Subject Headings list, and recognise the hard work of those who have made this possible. As champions of high-quality, useful metadata, we aim to highlight the value of specialist vocabularies produced with the aid of subject experts, and it is always a pleasure to see such projects come to fruition. Great work!”

Philip Keates and Martin Kelleher
(Co-editors of *Catalogue and Index*)
on behalf of the CILIP Metadata and Discovery Group
(https://www.cilip.org.uk/members/group_content_view.asp?group=201298&id=686430)

[1] Bina Vaidya, Gita Thapa, “Creation of Nepali subject headings with international standard : a project “, *Catalogue and Index* 202, pp. 48-51, and Bina Vaidya, “The Need to Develop a Collection of Subject Terms on Nepal’s Diversity”, *Catalogue and Index* 203, pp. 19-23.

Book Review: Edwards, Kimberley A. and Mackenzie, Tricia (eds.) (2021) *Telling the Technical Services Story: Communicating Value.* Chicago: ALA Editions. 240 p. $64.99. ISBN 978-0-8389-4946-7

The title of this collection of case-studies invites two inferences. First, "technical services," the common American term for back-office library processes, especially acquisitions, cataloguing, and e-resource management, suggests a US context. Secondly, "communicating value" implies a focus on advocacy for our sector of the profession. Only the first inference is correct. The editors note (p. ix) that all the contributors are "staff in research institutions across the United States," but they feel that the volume "can nevertheless provide a road map for an array of libraries." They are probably right; there is much here that could be of value to the UK's community of metadata librarians. Most chapters discuss work within the whole "technical services" department, but there are some with a narrower focus, including three with a heavy cataloguing/metadata element (all discussed below), and many of the issues touched upon are relevant to UK libraries.

Despite the claim that this book illustrates ways "to communicate the role [technical services departments'] work plays in supporting the mission of both their library and their larger institution" (p. ix), only a few of the case-studies truly have this kind of communication as their focus; in others the "communicating value" element is tangential; in still others it is barely there at all. This is not a major problem, but readers who obtained a copy based only on the title are likely to be disappointed. The real linking thread between the different chapters is much broader: communication within and by technical services departments. This more diffuse approach means that there are fewer synergies between the different chapters. As with many collections of essays, it felt at times a bit like reading a series of journal articles rather than a single volume. But the book contains many examples of good practice.

Several chapters focus on communication within or between back-office teams, describing the use of technology (Confluence in chapter 1, Basecamp in chapter 2, Trello in chapter 7) or techniques (retreats—i.e., away days—in chapter 3, a community of practice in chapter 6). The strongest of these chapters also argue persuasively for the importance of communication, such as when Faulkner and Sandford describe how good documentation fosters equity (p. 5): "documentation ensures equal access to critical information, mitigates incidental or intentional gatekeeping, creates transparency in local-decision making and cross-divisional communication."

Other contributions discuss communication with subject specialists or circulation desk staff, such as Mi, Pollock, and Falato's description of their provision of basic cataloguing and metadata classes to public-facing colleagues. Their project's aims include "that information could be exchanged, needs heard, and a common understanding established" (p. 82). That the intention was to develop genuine two-way communication is evinced by the survey they sent prospective attendees, asking them to choose the topics that would appeal to them from a list of possible options, such as how to read MARC records, how metadata impacts the OPAC, and what is linked data.

Towards the end of the book the focus shifts to communication with groups outside the library. Hargis and Novacescu describe the involvement of a cataloguer (Hargis) in marketing a library's resources. Initially seeing it as peripheral to her job, Hargis later came to view marketing as more integral, a way of demonstrating the library's value to users. The authors argue that cataloguers are often uniquely placed to promote a library's content, "through their knowledge of the value of a resource for a given audience" (p. 154). This resonated with me in my current work, where we have recently completed a project to catalogue the library's archival collections. There is nobody in the university who knows the archives, or understands their potential value, better than the people who did the cataloguing, and it is our responsibility to ensure that this value is communicated to users

Perhaps the most interesting chapter is Maggie Dull's description of the Metadata Outreach service at the University of Rochester. This service offers staff and students at the university the chance to consult with the library's metadata experts, and it arose out of a single interaction with an academic, whom the metadata team advised on a research data management query. I have known metadata librarians provide this kind of advice at other universities, but what is exceptional about Rochester's work is that they have built on this chance connection to provide a permanent service that is promoted throughout the university. Other examples of what Metadata Outreach has provided so far include advising the Rochester Digital Scholarship Lab about metadata schemas and standards, and consultation on data governance for the university's HR system. The success of such a service is largely dependent on how well it is promoted, and at Rochester it helps that both liaison librarians and the university's chief data officer refer colleagues to the service, and that the Metadata Outreach webpages are located prominently within the library's website.

Notwithstanding its slightly misleading title, this is, overall, a useful book. The contributions all have a practical focus, are clearly written, and cite the literature only when it is relevant. The volume is well edited and pleasingly presented. It is certainly worth consulting for anyone interested in how technical services departments do and should communicate.

[1] For the services offered by Metadata Outreach see https://www.library.rochester.edu/services/metadata-outreach/projects-and-services.

Catalogue & Index is electronically published by the Metadata & Discovery 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:

Philip Keates: p.keates@kingston.ac.uk

Martin Kelleher: martind@liverpool.ac.uk

Book reviews: Please contact the editors.

ISSN 2399-9667

MDG website: cilip.org.uk/mdg

MDG blog: http://catandindexgroup.wordpress.com/

MDG twitter: @CilipMDG

Tags from the MDG 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