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 2021, Issue 203

Welcome to C&I 203, which looks at some of the many collaborative initiatives that are taking place between, and beyond, libraries, both nationally and internationally.

In her first of two articles for this issue, Helen Williams shares her experiences contributing to the Metadata 2020 collaboration, and shows how working to improve large-scale issues can bring personal development benefits as well.

Martin Kelleher then sets out the case for establishing a NACO funnel, exploring at the benefits for institutions and the potential difficulties of implementation, and looking at how the final step to contribution might be taken.

Sarah Armitage and Heather Sherman of BDS walk us through the rationale behind their new Academic Library Licence (ALL), and answer questions on coverage, standards, and future-proofing.

Titia van der Werf-Davelaar looks at some of the key findings from the "Next Generation of Metadata" round table discussions organized by OCLC Research in March 2021. Prominent in the discussion were two national metadata programmes - one in France, one in the Netherlands - and there was also much interest in current efforts in the UK to better manage the metadata supply chain.

In a follow-up to her and Gita Thapa's fascinating article from last issue, Bina Vaidya gives us a snapshot of headings for languages, ethnic groups, costumes, festivals, and musical instruments from across Nepal, highlighting the great cultural diversity of her country, and showing how important subject schemes created with local knowledge and expertise are.

CILIP SPECIAL INTEREST GROUP

OPEN ACCESS

Contents

One of the major efforts being undertaken to understand and manage the UK metadata supply chain is the BIC Metadata Capabilities Directory, a resource which will allow BIC members to better understand the metadata needs and limitations of those they deal with. Project lead Peter Mathews takes us through the development and workings of the MCD.

We finish with Helen William's second article, in which she describes the progress LSE have made in their recent work with Wikidata. It's a piece full of useful links, examples, and practical advice, and hopefully it will inspire others to begin their own collaborative journeys with Wikidata.

Enjoy - and keep collaborating!

Martin Kelleher: martind@liverpool.ac.uk

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

Working with Metadata 2020

Helen Williams, Metadata Manager, LSE Library

The LSE Library Strategy recognises that 'Library staff have a unique combination of professional knowledge, skills and experience' and inspires us to 'act as leaders in our professional networks in the UK and internationally, ensuring that our involvement contributes to the School's mission.'[1]

Having an employer who encourages and supports external involvement in the wider profession can enrich the careers of those who take hold of this, as well as benefitting their institution, and my own working life has been greatly enhanced by the opportunities to get involved in work beyond the walls of the Library.

My first experience of working in a group outside my workplace was with the CIG (now MDG) committee. This gave me an excellent opportunity to work alongside metadata specialists at other institutions, learning from their professional expertise, as well as sharing what I had learned and developing my own skills as I took part in the discussions and work of the committee. Those relationships continue to this day, enhancing my professional life (and extending to friendships beyond it). By the end of 2016 I had been on the committee for seven years, which was the maximum term of service allowed by CILIP without a break. Since joining the committee I had moved roles within the LSE Library, and now had a broader remit managing its Metadata team, but I was still keen to make time for professional involvement to contribute to the sector, and to enrich my day-to-day work.

In early 2018 I became aware of the work of Metadata 2020, 'a collaboration that advocates richer, connected, and reusable, open metadata for all research outputs, which will advance scholarly pursuits for the benefit of society.'[2] The collaboration involves librarians, funders, researchers, publishers, aggregators and service providers and opened up the opportunity to work alongside a global community of practitioners from across the scholarly communications ecosystem. I volunteered to join two of the six project groups focusing on Researcher Communications and on Shared Best Practice and Principles.[3] Initially I took part in online meetings and contributed to some offline work. Then in October 2018 a two day workshop was convened for participants across all six projects. I was eager to attend and become more involved but I had some degree of uncertainty as to whether I would be able to make a valuable contribution once I saw that most of the attendees came from the other professional groupings. These fears were alleviated on the day and I found that my enthusiasm for the topic, and desire to get to know fellow attendees were shared by other participants who, just like the library community, were friendly, communicative and willing to bring me up to speed on any sector specific gaps in my knowledge. Many of the topics and issues that came up in discussion were familiar ones, particularly as I work with research outputs metadata as well as the more traditional forms of library metadata, and I realised that I was able to contribute experiences from my library perspective.

It is not unusual for opportunities for further involvement with community engagement to come with a heightened sense of imposter syndrome, and this was my experience when during an online meeting the week before the workshop I was asked to deliver the project update for the Shared Best Practice and Principles group, and to chair an interactive session. Despite some trepidation this was an excellent opportunity for a new professional challenge as well as the chance for some personal development that would require me to step out of my comfort zone. The project update was less daunting because as a project participant I was familiar with our work, and was able to prepare an update in advance, which one of the project leads, in America, kindly fleshed out for me. Chairing a session, however, felt significantly more challenging. As a project team we had been working, in online meetings and offline work, on a set of metadata best practices and principles. These had been discussed at a workshop in New York, which took place the week before our London workshop, and an online call (in the pre-pandemic days, before this became our normal way of working!) to one of the project leads enabled me to be brought up to date with what had been discussed and agreed, and ascertain what needed to be achieved at the London meeting. Chairing sessions outside my workplace was a new experience for me, using a different skill set from my day-to-day metadata work in order to respond to the unexpected, follow fast discussion and debate, and know when to follow diversions and when to bring the conversation back on track.

It was my responsibility to capture discussions and report back to the leads of the Shared Best Practice and Principles project, but mindful of the fact it could be challenging to facilitate discussion at the same time as taking comprehensive notes, the workshop convenor kindly agreed also to take notes of the session and these were very helpful in refreshing my memory after the event when I came to write up the discussion. Initially we discussed and captured ideas, but as the conversation grew and new ideas formed it was helpful to capture these visually. We therefore used a whiteboard so that everyone could see outputs and we could jump more easily between principles and points of best practice, capturing relationships where required and editing our work as our thinking changed and developed.

Once a completed draft of the best practices and principles had been finalised by the project group I responded to a call to co-write an article about this community-based effort. This gave me the opportunity to work closely with the project convenor and four core team members in the United States to share the methodology used to create the principles, connect them to existing work, outline their use and suggest how they might serve as a foundation for future activity. This is the second time in my career that I have co-authored an article with external colleagues and it is an enriching experience to observe, learn from and work alongside experts in closely related professions.

At the beginning of the year 2021, Metadata 2020 relaunched their website with the purpose of supporting ongoing community activity[4] and ownership. This encouraged direct action, with 'things you can do to promote and support richer, open metadata'[5] and promoted learning through sharing the outputs created by the Metadata 2020 community[6].

Overall, it has been an enriching and enjoyable experience to work with a broad and diverse community so relevant to my library metadata focus and I am delighted to have been part of an initiative promoting the value and necessity of metadata in connecting research for the benefit of society. We should never underestimate the value of our metadata skills or their contribution to the world around us. 'From a societal context, open metadata is a critical component of addressing some of the world's greatest challenges.'[7] I would encourage our metadata community to 'think big', grow confident in our abilities to contribute to and enrich society through our metadata, and to be involved wherever we can be.

References

[1]LSE Library (2016), *LSE Library Strategy,* https://www.lse.ac.uk/library/about/lse-library-strategy (Accessed 27/4/2021)
[2]Metadata 2020 (2020), *About Metadata 2020,* http://www.metadata2020.org/about/ (Accessed 27/4/2021)
[3]Metadata 2020 (2021), *Phase 2: Projects,* https://metadata2020.org/learn-more/projects/ (Accessed 27/4/2021)
[4]Kaiser, Kathryn, Kemp, Jennifer, Paglione, Laura, Ratner, Howard, Schott, David and Williams, Helen K. R. (2020). Methods & proposal for metadata guiding principles for scholarly communications. Research Ideas and Outcomes, 6. ISSN 2367-7163, https://riojournal.com/article/53916/ (Accessed 27/4/2021)
[5]Metadata 2020 (2021), *Do More,* https://metadata2020.org/do-more/ (Accessed 27/4/2021)
[6]Metadata 2020 (2021), *Phase 3: Outcomes,* https://metadata2020.org/learn-more/outcomes/ (Accessed 27/4/2021)
[7]Metadata 2020 (2021), *Spread the word about metadata* https://metadata2020.org/do-more/spread-the-word/ (Accessed 27/4/2021)

Establishing a national funnel to the Library of Congress Name Authority File: Making it happen
A call to action by **Martin Kelleher**, Metadata Manager, University of Liverpool on behalf of the CILIP Metadata & Discovery Group

Abstract

Having recently transferred to NACO (Name Authority Co-Operative) headings from a local standard, investigations into effective authority control at the University of Liverpool seem to indicate the benefit of being able to contribute to the central NACO file to best manage local management. Individual contribution requires significant quota fulfilment which encourages the investigation of funnel contribution, currently unutilized in the UK metadata community. Investigation of this possibility have led to the discovery of the significantly improved feasibility of this approach, not least from changes in management of funnel contribution centrally by the PCC (Program for Cooperative Cataloging),[1] which encourage the establishment of a UK funnel as being a desirable, substantial and achievable proposition. This paper discusses progress towards the creation of a UK funnel.

Background – The Liverpool Perspective

At the University of Liverpool, we'd recently changed from local authority control to NACO.[2] We'd previously only used local authority headings where necessary for indicating preferred versions of names where an individual or organisation was known by several (e.g. spelt different ways in different languages, or with different surnames based on changes of marital status etc.) when we were using our own authority file. We had also not used the authority module available on our system (Sierra) to automate control, and indeed weren't even aware that this function was available. Activating this function was intended as part of the plan in transferring to NACO.[3]

When we transferred to NACO, it seemed to me to be highly desirable to be able to create and edit NACO entries, since I could see too much room for logistical conflict if we tried to manage headings using local entries, in that there would always be a risk of local amendments being conflicted with core updates by NACO, which is updated periodically. Additionally, I was keen to investigate the development of practice of adding extra entries in authority records to facilitate automatic control, something which may be more of an issue in a mixed source data economy *outside* of the US, where NACO entries, held in the NAF are less universal in application.

Library of Congresses NAF (Name Authority File),[4] is the primary source of authority data for metadata across the English speaking world and beyond. Based as so much is, in Library of Congress, and the PCC project, the Name Authority File is expanded by contributors, who undergo a training course in authority record creation, with records being processed through institutional or regional 'funnels', the former of which are required to create a certain number of headings per year.

As an initial investigation, I started going through OCLC and PCC pages,[5] and discovered the requirements of high production for contributors directly contributing.[5] I became keen to investigate hub access, which dispersed this quota amongst contributors but it seemed there were no hub projects active in the UK. Investigations indicated that others had found this to be a far more accessible and useful approach to contribution.[7, 8]

The idea of organizing national involvement with the Library of Congress Co-operative Cataloguing Project is not new, but through my investigations I believe it is finally looking to be a feasible reality.

Background – the history of UK NACO proposals

Originally posited by Deborah Lee as early as 2012,[9] recently supported by WHELF (The Wales Higher Education Libraries Forum),[10] it was raised again in the first CILIP MDG committee meeting I attended (in December 2019), not by myself but by Paul Shackleton of OCLC based on my enquiry. It was subsequently

raised by the Chair, Jane Daniels, in the March 2020 meeting, which in my ignorance I didn't realise was based on WHELFs earlier investigation, of which Jane was also a member. I henceforth volunteered to lead on the project, only after further corresponding with Jane realising that I'd essentially poached a topic that was already being investigated by WHELF!

The UK NACO landscape

Jane encouraged me to investigate the UK contribution network, which I undertook primarily by corresponding with staff at the British Library, possibly with an eye to a co-operative approach. The British Library indicated however that they were direct contributors, as were all the UK Legal Deposit Libraries, unsurprisingly given that all such would presumably be able to fulfil the quota required by PCC for such. Alan Danskin was, however, in the process of investigating a SACO FAST funnel, indicating that here at least there were similar actions in the same direction by the British Library.

Why it is the right time to join NACO

As my investigations continued, especially as I undertook the training in September 2020, it became clear that there had been a number of changes in PCC policy and practice which made the organisation of a national Funnel more appealing, both on a local and national basis. I also communicated with Frank Koper, the Funnel Co-ordinator for the Netherlands Funnel, who indicated he'd found the implementation of the funnel to be more reasonably manageable.

The costs of training are….gone

One of the primary obstacles to widespread uptake of the project is cost both in terms of time and money to undertake the training. Recent changes in the way the training is delivered should remove much of the obstacles to uptake. My training, and likely much of the training, is now delivered online, to an increasingly international user base. This reduces financial cost to essentially zero, dramatically reduced from previous arrangements of having to arrange accommodation for trainers and any nonlocal trainees, plus travel for the former (from the US).

The Logistics are better

This leads to knock-on benefits in terms of the logistics in terms of organizing the training, largely reduced again by the convenience of this switch to web training, all of which is increasingly firmed up by PCC and particularly Paul Frank himself. Even better, the time management element of training has also improved.

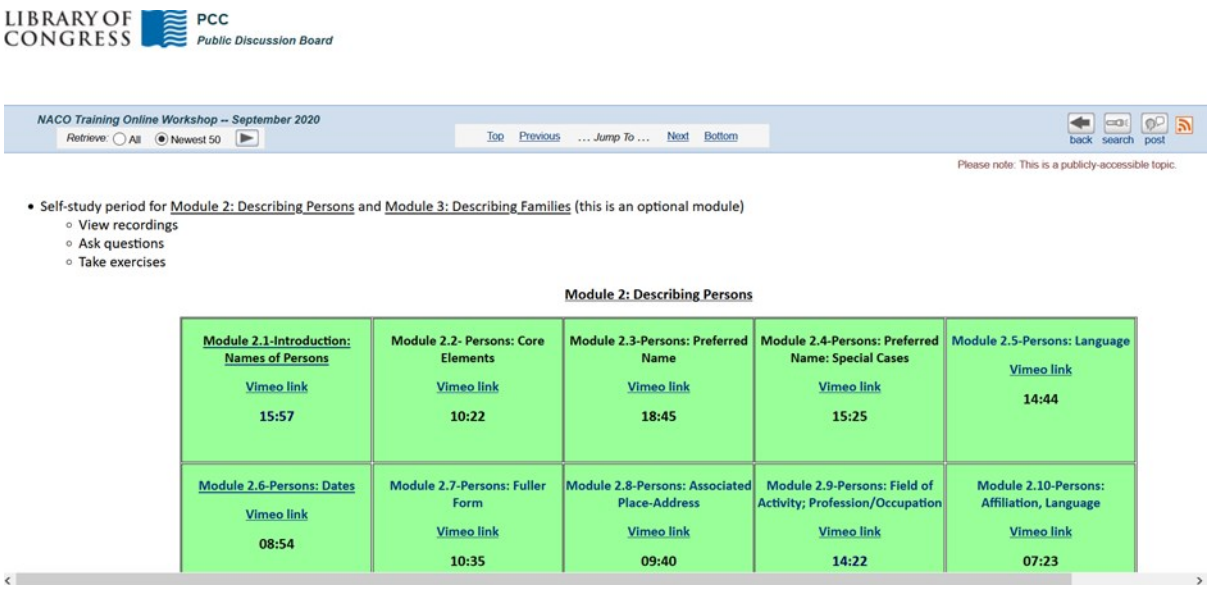


Figure 1. A sample of the training workshop

Time commitments are reduced and more flexible

Having undertaken the training in September 2020, it became clear that there was an understanding that the training would not necessarily be undertaken by everyone within the overall time period of the sessions, spread over 2 weeks, since much of the material was pre-recorded, making the learning process considerably less time sensitive.

This was good news to me, since a significant workload and programme of other simultaneous projects made full time engagement in training impractical, and this flexibility allowed me to spread the whole thing including practice authority creation over a period more like 2 months! Even better, however, is the news that, following on from a few comments made during the training, I discussed the matter with the head of the PCC project Paul Frank, and we agreed a reduced portion of the programme could be engaged for those interested in creating only certain kinds of records. Although I undertook the whole training course, I am unlikely to engage, in the short term at least, in Authority data for locations or uniform titles, or family names for that matter, and I suspect others are likely to also not require these or perhaps other areas, so the whole programme becomes considerably more flexible, less time consuming, and consequentially significantly more feasible to a greater number of institutions to engage in.

Ongoing requirements are reduced to… nil

Formally, there is an institutional requirement for contributing institutions to create a certain number of headings, which is formally spread out across a funnel. During the training I undertook, however, it was explained that funnel requirements have effectively been reduced to nil, presumably because contribution via a funnel is found to be more desirable for PCC management, and to encourage wider participation in contribution.

The added value of NACO training

Regardless of the requirement for NACO contribution, NACO training offers insight into the functional nature of authority headings, as well as constantly referring back to RDA standards referring to the same. As a library only recently formally switching to RDA as local practice, we struggled to find effective practical training in RDA, and the extent to which NACO training provides direct practical understanding of RDA and its practical application in NACO has acted as a substantial element in my own education in this new standard. I would suggest that this would similarly act to benefit other metadata librarians in understanding the dominant standard in modern metadata in application.

Ongoing costs are…. reasonable? And may become more reasonable?
The funnels are maintained by OCLC, and having discussed the issue with OCLC, they have indicated that they require a cataloguing subscription to access funnels (and the systems required to access it), and that subscription has a base institutional rate of £1,175. However, even this might be something that is looking likely to improve nationally, following recent proposals in terms of a JISC based national subscription to OCLC, but nonetheless offers a basic figure for acquiring access, along with other benefits.

While it became clear that NACO contribution was becoming significantly easier to establish and support, the value of authority management also became increasingly clear as I investigated it, as was the likelihood of this value increasing over time.

The value of being a NACO contributor

Empowerment to input into global dataset

The value of having a UK funnel would largely to be in order to add entries into NACO, or modify them not only for local purposes but also for contribution to a larger knowledge base.

Empowerment to maintain local control via global dataset

As mentioned earlier, I wanted to be able to add and edit authority records without those entries removed, or modify them not only for local purposes but also for contribution to a larger knowledge base.

Empowerment to maintain local control via global dataset

As mentioned earlier, I wanted to be able to add and edit authority records without those entries removed, or relevant changes deleted, or duplicate entries from NACO being loaded and duplicating locally created data whenever we reload and refresh the authority file. By inputting changes required for local management into the global dataset, this liability is essentially bypassed.

Foundations for the future

On a larger and more progressive scale, increased use of authority data is very much the kind of direction RDA moves towards. Authority records are URI based linked data, and the collaborative approach involved in it's management and functionality takes us in the direction that libraries and the world of information management is going in.

National benefits

On a national scale, the UK is underrepresented in many areas of the NAF, something which more widespread involvement would correct. The benefits of this could be significant. The use of consistent entries for UK authors would improve discoverability of the same in a way that is not perhaps currently catered for, and may well also have related benefits for statistical purposes, in that having more managed entries for UK authors could enable more effective and accurate statistics regarding the use of their materials.

Promotion and identification of locally relevant persons

When the whole project came up at Liverpool, it was mentioned that there were local authorities for persons relevant to local special collections within the library. In communication with Debbie Lee, who was investigating UK funnel contribution in 2012-2014, the relevance to Academic REF management was also referred to, and of course there is value in managing locally relevant headings which for this and other purposes benefit from improved ability to manage authorities on a global scale effectively. For libraries not maintaining their own authority file but relying on NACO or another authority file really requires contributory rights to allow this to be employed effectively.

Questions?

There remain some questions that are not fully resolved however:

Cost

Although OCLC rates may seem reasonable to those such as Liverpool University with widespread use of datasets, individual downloads, and not least for our initial conversion to NACO, if a subscription is not held by an institution, especially a smaller institution, it may still be a barrier to membership. There is the hope of the UK wide license to access OCLC currently in consultation with JISC, but this is yet to be confirmed or agreed. Discussions with Paul Frank and Debbie Lee, however, have indicated that OCLC are only necessarily involved for registration so, especially if the JISC deal mentioned earlier is not undertaken, the possibility and process of nonsubscriber registration and modes of access outside of Connexion Client should be established. At Liverpool University I'm in the process of setting up access via Innovatives Skyriver, in order to embed authority contribution in a more efficient process, so there is at least one other approach available.

NACO or NACO/SACO?

The British Library are planning their own FAST funnel,[11] which they intend to open up to the community, and it was mentioned that perhaps this could be used to allow standard LCSH submission also, as an option to the funnel I was in the process of setting up for the MDG. Uncertain as to the likely uptake for LCSH creation, I was considering in the first instance only creating a NACO funnel, and hoping to leave what I was expecting to be a smaller interest in SACO to occasional input through the British Library FAST funnel.

However, times are changing and there has recently been an increased interest in subject headings, led by developments such as the "Change the Subject"[12] Campaign in the US and transatlantic Cataloguing Ethics[13] project. We've recently also run a subject indexing issue for Catalogue &Index,[14] and the interest in developing subject indexing, including LCSH on these and a number of other areas has prompted me to consider, should we be looking to ultimately facilitate both NACO and SACO for the funnel after all? Whether a single or dual funnel should ultimately be established, and whether established at the onset of the project, are decisions to be made based on interest and the needs of the metadata community.

How to launch a UK Funnel

I was going to set up an individual webinar to promote and discuss the establishment of a funnel, but with all the various other initiatives going on, not least the OCLC JISC license, I'm wary it might be overlooked, so now I'm thinking of launching it at the MDG mini-conference we're hoping to have later in the year, but that would mean missing the next wave of training in September. I'm not entirely keen on leaving it until then to start the funnel, so now I'm beginning to think it might be worthwhile kickstarting the funnel with a few early adopters.

The stumbling plod of Progress

I think another stumbling block is also the technological aspect of authority control. Sierra is not the best at facilitating automatic control, and discussing the issue with Northern Collaboration peers indicated that Alma may also be overly reliant on the same kind of manual intervention to check accuracy of changes. Differences in case, punctuation and diacritics are all obstacles to automatic process in Sierra, requiring far too much manual intervention to make the process as efficient as I would hope, and whether the same issues are present and obstacles to automation in Alma is something I'd be concerned about.

A second concern is that as an essentially Linked Data format, authority records do not benefit from anything like the full functional support they could benefit from to fulfil this function... Lobbying for more facilitated, more automated, and less sensitive automatic authority control systems would be a long-term goal, which itself would be supported by more overt active employment of authority control.

Neither of these elements are issues that should impede the establishment of a funnel, and indeed the establishment of a national funnel would establish an organizational structure as of itself which would strengthen the UK metadata community in dealing with these issues for the empowerment of libraries to facilitate these developing technologies which currently are under-facilitated by prominent library systems.

Just a little bit longer? Or not?

The terms are good, and the time is right, I believe the foundation of a UK library hub are within reasonable grasp, all expressions of interest particularly regarding early adoption are invited to contact myself if you are interested in an initial meeting to consider training and involvement in a national funnel to the NAF, and take part in taking your library and national metadata as a whole into a richer, more empowered position.

Please email martind@liverpool.ac.uk to indicate further interest, or to register for training to become an earlier adopter.

References

[1]PCC homepage https://www.loc.gov/aba/pcc/ (accessed 17/6/2021)
[2]Kelleher, Martin, 2020. Going to the Co-Op: converting authority headings from local to global headings using datafrom OCLC Worldshare at the University of Liverpool. *Catalogue & Index*, 201, pp. 29-43
[3]NACO homepage https://www.loc.gov/aba/pcc/naco/ (accessed 17/6/2021)
[4]NAF homepage https://id.loc.gov/authorities/names.html (accessed 17/6/2021)
[5]OCLC NACO help pages https://help.oclc.org/Metadata_Services/Connexion/Troubleshooting/How_do_I_become_a_NACO_participant?sl=en (accessed 11/6/2021)
[6]Frequently Asked Questions about Joining the NACO Program – 20. Is there a minimum number of headings we must submit annually or monthly? https://www.loc.gov/aba/pcc/naco/nacoprogfaq.html#20 (accessed 11/6/2021)
[7]Lisius, Peter H., 2008. One Cataloger's NACO Participation: Comparing Funnel Participation and PCC NACO Classroom Training. *TechKNOW*, 14 (3), pp. 1-5
[8]Jones, Julene L., 2015. Many Hands Make Light Work: Name Authority Cooperative Program (NACO) Training at University of Kentucky Libraries. *UKnowledge Library Presentation* https://core.ac.uk/download/pdf/232566281.pdf (accessed 17/6/2021)
[9]Lee, Deborah, 2012. In the name of collaboration. *Catalogue & Index*, 169, pp. 30-36 https://cdn.ymaws.com/www.cilip.org.uk/resource/collection/F71F19C3-49CF-462D-8165-B07967EE07F0/Catalogue_and_Index_issue_169,_December_2012.pdf
[10]WHELF action plan 2020/2021 https://whelf.ac.uk/wp-content/uploads/2020/11/WHELF-Action-Plan-202021-priorities-1-1.pdf (accessed 11/6/2021)
[11]Danskin, Alan, 2018. FAST forward: Sustaining and extending subject indexing at the British Library. *Catalogue & Index* , 192, pp3-8
[12]Broadley Sawyer et al., 2019. Change the Subject. https://collections.dartmouth.edu/archive/object/change-subject/change-subject-film (accessed 18/6/2021)
[13]Cataloging Ethics Steering Committee. Cataloguing ethics https://sites.google.com/view/cataloging-ethics/home (accessed 18/6/2021)
[14]Catalogue & Index, 202.

The BDS Academic Licence (ALL)

Sarah Armitage, Director of Library Sales, Bibliographic Data Services Limited
Heather Sherman, Director of Academic Library Operations, Bibliographic Data Services Limited

Introduction

Five years ago, Jisc started talking to the HE sector about developing a national bibliographic knowledgebase,[1] which would have as one of its aims the improvement of metadata accuracy, effectiveness and standards. Since then, BDS has been in close discussion with Jisc, stock suppliers, and key stakeholders about how our services and knowledge could be used to help improve metadata services to libraries, which would free up cataloguers to use their expertise where they can add most value to their institutions. This led us to develop the BDS Academic Library Licence (ALL) which offers a more collaborative approach to metadata provision and sharing. The emphasis is on creating a record once for routine acquisition titles, which is then available across the sector to all participating libraries, stock suppliers, eBook aggregators and publishers, thus driving efficiencies for everyone, whilst also creating time for cataloguers to work on unique and specialist services, collections and data, and improving discoverability for the end user.

Who are BDS?

BDS are Bibliographic Data Services and much of the high quality data in UK library systems originates with us. A quick check of the 040 for StDuBDS will show how many records in your LSP have input from BDS. Our cataloguers have supplied the British Library with CiP (pre-publication) data which maintains the British National Biography (BNB) for over 20 years. BDS's Cataloguers are metadata professionals, expert in generating accurate and consistently high-quality metadata, and sit on a wide range of national and international standards committees.

How is the ALL different from how I obtain MARC records now?

MARC records are currently provided in a mixed economy with data being created by BDS and also by library stock suppliers, publishers and aggregators. This results in duplication of effort, and a huge variation in quality between the records being provided which can lead to Cataloguers not having confidence in the metadata which is being loaded into their library systems, a finding which was highlighted in Jisc's Remodeling the UK Data Marketplace report,[2] and in NAG's Quality of Shelf Ready metadata report.[3] Currently, records are created pre-publication and often not upgraded with the book in hand. The ALL seeks to address these issues by creating full, book in hand records either on or before publication for the majority of titles, thus ensuring that a reliable record is provided once which can be accepted into the LSP without the need for local checking and amendments, creating efficiencies across the entire supply chain.

As a Cataloguer, how will the ALL help me?

BDS recognises the huge benefits that Cataloguers offer to their institutions in creating and maintaining metadata to ensure that resources can be found and used as easily as possible. By meeting one of the aims of Plan M[4] to regard metadata for routine acquisitions as a solved issue, the ALL will free up Cataloguers to use their specialist knowledge and expertise to provide metadata and services that cannot be outsourced. Universities are never short of projects that require the skills of Cataloguers: this could be maintaining Reading List systems, providing advice and guidance on metadata management for research repositories, or cataloguing the over 13 million volumes in Special Collections that the RLUK's Hidden Collection report estimates were uncatalogued in 2012.[5]

So does the ALL mean that I would no longer have to spend time searching for, downloading and copy cataloguing records?

The ALL provides full, Book-in Hand records for new additions to stock for all English language print, ebook or

open access titles published after 1st January 2010 or for which BDS already has a full record. The 2010 date was chosen after close analysis of HE purchasing over the last three years, which showed that the scope of the licence would cover the vast majority of the records required by UK and Irish universities. These records will be provided in response to orders being placed and so will be supplied as a matter of course to accompany the delivery of the resource.

What will the content of records supplied under the ALL be?

In preparation for the ALL, BDS has worked closely with the SUPC to ensure that the records which are provided will meet the requirements of the new stock supply agreement, and has also ensured that records created under the ALL very closely meet the recommendations of NAG's "Quality of Shelf-Ready Metadata" report. In response to the NAG report, BDS is working to include additional identifiers, such as the British Library National bibliography number in MARC21 field 015 and the NACO control fields for name headings. The presence of this information in BDS's MARC21 records will improve record matching and de-duplication efforts, and will improve authority control, both of which are hugely beneficial to shared databases.

Will records supplied under the ALL be RDA compliant?

BDS was the first agency in the world to implement cataloguing using the RDA standard, making the switch in January 2013. For those records predating 2013, a conversion program can be employed to apply RDA principles retrospectively to AACR2 records, so current standards can be maintained even if older material is purchased by a library.

What about quality control, will I have to check records supplied under the ALL?

We are conscious that one of the frustrations that was highlighted in the Jisc Remodelling the UK Data Marketplace report was that Cataloguers were spending time amending shelf ready records. BDS records are created by our team of highly qualified and experienced Cataloguers and adhere to all current agreed library standards, meaning that quality checking of records provided under the ALL by individual institutions will no longer be needed.

Will records for ebooks which I might be purchasing from publishers or aggregators be covered under the ALL?

BDS will provide high quality MARC records in response to orders for ebook titles in the same manner as orders for print books. BDS is working closely with ebook publishers and aggregators to ensure that high quality ebook records are created and made available as early in the publishing cycle as possible. ALL holders can also use the BDS database to download good quality records for ebooks already in their catalogue.

Can the ALL make my workflows more efficient?

We hope so! The ALL mirrors the model of the BDS Public Library Licence which is used by every local authority in the UK and which has enabled BDS to work with public library LMS suppliers to devise innovative methods for the seamless, hands off delivery of data. BDS is working with the academic LSP suppliers to offer the same workflows to ALL holders whereby MARC records are automatically delivered in response to orders without the need for any intervention by library staff.

What about the future, BIBFrame and Linked Open Data?

BDS is, and has always been, at the forefront of emerging library standards and new developments and is involved in all aspects of developing and maintaining bibliographic standards in the UK. For example, BDS was the first organisation in the world to implement cataloguing using the RDA standard. We ensure that we keep abreast of developments in library standards and cataloguing and also in the wider book trade through our involvement in many varied national and international committees and working groups.

The development of the MARC21 standard continues to look at ways to address the increasing preference for expressing data in the context of Linked Open Data, and with BDS staff involved in national committees which provide feedback to the MARC Advisory Committee, we are well placed to help librarians optimise their systems and to achieve the full range of benefits going forward. The Library of Congress' new bibliographic format, BIBFrame, will continue to be mapped to MARC21 for exchange purposes and so does not replace MARC21.

Holders of the ALL will be invited by BDS to participate in a User Group to shape the future of the Academic Library Licence and to ensure that it meets the needs of the UK HE library community, both now and in the future. A dedicated helpdesk for Licence holders will provide direct access for Cataloguers in academic libraries to BDS's expert team of practitioners who will be available to answer questions and help with cataloguing queries.

BDS looks forward to being able to work collaboratively with its academic library customers to provide a high quality, fit for purpose metadata service which will provide efficiencies both now and into the future.

For more information about the BDS Academic Library Licence, please contact Heather Sherman, Director of Academic Library Operations at BDS at heather.sherman@bdslive.com on 07711 378086.

References

[1]https://libraryservices.jiscinvolve.org/wp/2016/07/the-national-bibliographic-knowledgebase/ accessed 24/5/21
[2]https://libraryservices.jiscinvolve.org/wp/files/2020/11/Remodelling-the-UK-Library-Data-Marketplace-Full-slidedeck-PPT-Accessibility-Checked.pdf accessed 24/5/21
[3]https://nag.org.uk/wp-content/uploads/2021/04/NAG-Quality-of-Shelf-Ready-Metadata-Survey-Analysis-and-Recommendations_2021Corrected.pdf accessed 24/5/21
[4]https://libraryservices.jiscinvolve.org/wp/2019/12/plan-m/ accessed 24/5/21
[5]https://www.rluk.ac.uk/rluk-hidden-collections-report/ accessed 24/5/21

Mobilizing community to transition to the next generation of metadata

Titia van der Werf-Davelaar, Senior Program Officer, Research, OCLC B.V. EMEA

Abstract

This article highlights community and stakeholder mobilization initiatives in the library and heritage sectors that help in the transition to the next generation of metadata. We draw from the Next Generation of Metadata round table discussions organized by OCLC Research in March 2021. In these discussions, we saw next generation metadata mobilization taking place along two trajectories: (1) transforming and publishing institutionally sourced metadata and (2) improving metadata already in the supply chain. The article provides context and scope from the round table conversations and highlights national initiatives taking place in both mobilization areas. The article then discusses the challenge of managing at multiple scales, as efforts of local, national and global scale gear up to connect with each other.

Introduction

OCLC Research facilitated eight virtual round table discussions[1] with metadata experts on the topic of transitioning to the next generation of metadata, throughout the month of March 2021. In total, 86 participants from 71 different institutions – mostly university libraries, bibliographic agencies and national libraries – contributed to the sessions. Sessions were held in six different European languages so that participants could comfortably contribute to the discussion.

The main discussion question, "*How do we make the transition to the next generation of metadata happen at the **right scale** and in a **sustainable manner**, building an **interconnected ecosystem**, not a garden of silos?"* implies the need for intentional coordinated action towards achieving a shared vision of this future ecosystem. In the succeeding sections, we first provide the context of these conversations and then we present coordination and collaboration approaches pertinent to achieving transition.

Context

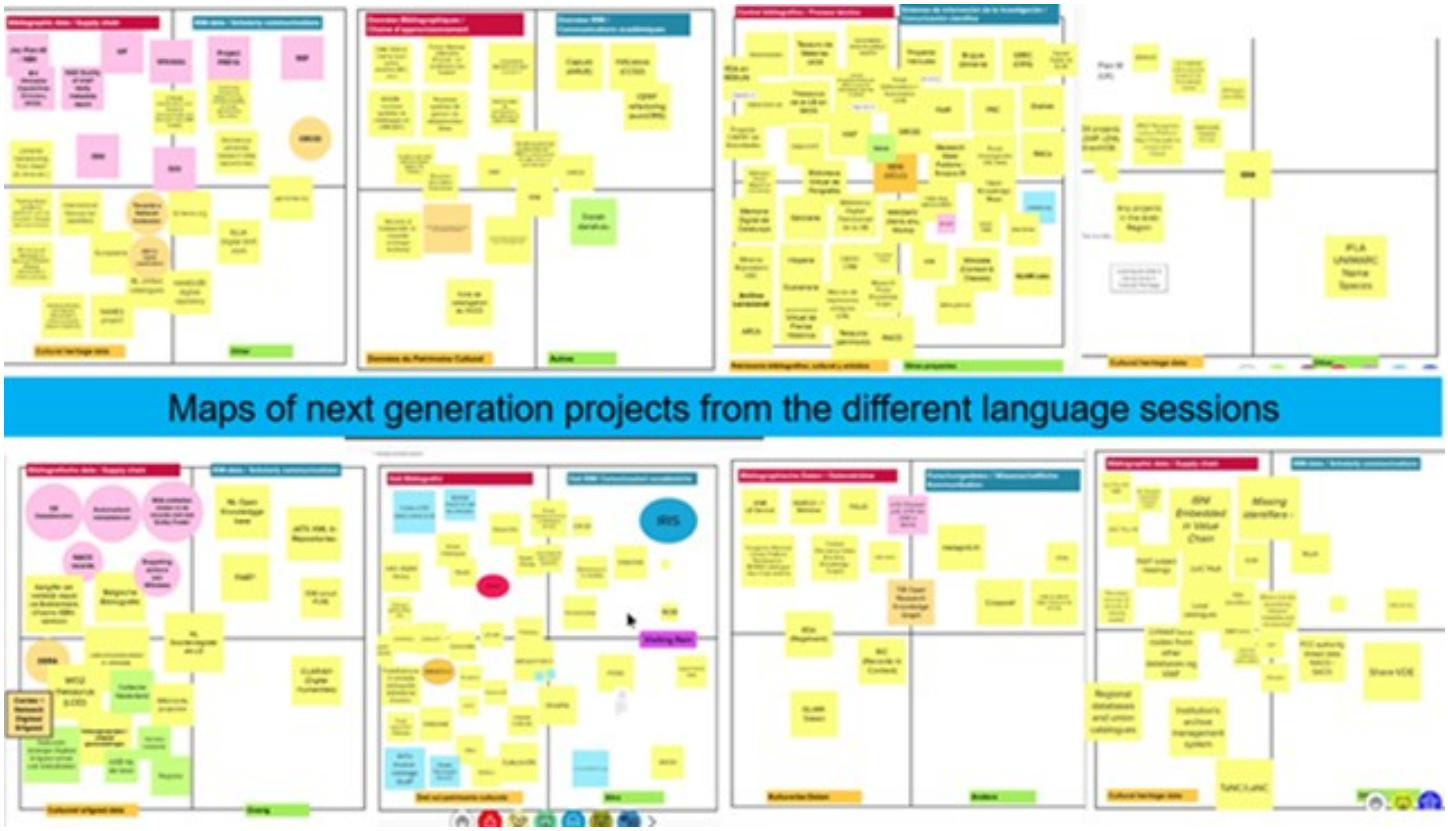
Karen Smith-Yoshimura's report, "Transitioning to the Next Generation of Metadata"[2] served as background reading and inspiration for the sessions. This report consolidates six years (2015-2020) of discussions on topics covered by the OCLC Research Library Partnership Metadata Managers Focus Group[3] about the future of metadata and the changes affecting their core areas of work: Creating and managing metadata. The changes taking place include the transition to linked data and the adoption of identifiers such as URIs (Universal Resource Identifiers), the move of metadata creation upstream in the data supply chain, and the evolution of metadata as a service. These changes open opportunities for libraries to engage in new areas where metadata is becoming key, such as Research Data Management (RDM) and Research Information Management (RIM). These changes are also key to achieving important library goals, such as support for multilingualism – which in turn, is closely connected to EDI (Equity, Diversity and Inclusion) principles.[4]

Working from the premise that the "what" and the "why" of the future of metadata operations are well understood, the round table sessions aimed to address the practical concern of the "how to" and "with whom" – a concern aptly articulated by Jane Daniels, moderator of the CILIP Conference 2020 session where Karen presented her report:[5]

"*I have read the report right through, in fact I have read quite a lot of it twice now. And every time I read it, I think – yes, I can see the direction we should be going in. I suppose the question is, how do we mobilize, probably at an international scale, to make sure that what you are listing, or what you are putting in the report actually comes about*".

To get the conversation started we asked attendees to map next generation metadata projects they were aware of on a 2×2 matrix characterizing three different application areas: 1) bibliographic data and the supply chain, 2) cultural heritage data and 3) metadata to support RIM and scholarly communications.

The collage of the eight maps that came out of the round tables shows a large number of cultural heritage data projects and bibliographic data projects and few RIM projects, reflecting the focus and expertise of the attendees. Clusters of post-it notes naming persistent identifiers (PIDs) – such as ISNI, ORCID, VIAF, DOI – featured prominently in all the maps, showing their value to participants.



The conversation then took deep dives into some of the initiatives, projects and services in order to analyse what it takes to make the transition to next generation metadata ***scalable and sustainable*** and to build an ***interoperable and interconnected ecosystem***.

In the next section we present coordination and collaboration approaches that stood out because of their intent to mobilize communities and stakeholders in bringing about this transition.

Approaches geared toward mobilizing stakeholders and communities

Transforming and publishing metadata

From the round table sessions, we learned that libraries are deeply invested in transforming and publishing their authority files (both the name authorities and the subject headings), in order to leverage them. We heard that open government data policies and national interoperability frameworks are strong drivers for this focus. National libraries have been pioneers in the field, publishing their national bibliographies and authority files as linked data in the course of the past decade. Whilst many libraries utilize the national authority file maintained by their country's national library or a large international authority files - such as the Library of Congress Name Authority File (LC/NACO Authority File) – others maintain their own local authority files. Many institutions with major cultural heritage or academic collections are also publishing or planning to publish their local authority files and bibliographic data as linked data. They mostly act individually. As a result, their approaches and implementations tend to differ, leading to a range of idiosyncratic solutions and much duplication of effort. This problem was identified during the round tables. Innovative approaches to create opportunity for more concerted effort are still sparse, however, two initiatives were discussed - one in France and one in the Netherlands – that provide interesting models for such concerted effort at a country level.

In France the collaboration between the two main bibliographic agencies – the Bibliothèque nationale de France (BnF), and the Bibliographic Agency for Higher Education (Agence bibliographique de l'enseignement supérieur, Abes) – is seen as a breakthrough needed to lead the bibliographic transition in France. BnF and Abes were enticed to collaborate because of the harder delineation between metadata production and release on the one hand and metadata management and (re-) use on the other, instigated by the open government data policy. The aim of the collaboration is to maximize the benefits of centralization, normalization, and efficiency at production and publication time. The two agencies decided to build and populate the French National Entity File (NEF) and to produce bibliographic data following agreed standards and practices, based on RDA and IFLA models. To involve the French library and archive community they launched the national Bibliographic Transition Program,[6] which, in their own words, *"aims to support all stakeholders in France through the drastic changes induced by such a major normative and technological evolution*." The program encompasses the massive processing, refactoring, and publication of the French legacy heritage data, the country-wide adoption of the new French cataloguing code that is in development, consultation with providers of library systems and software solutions, and professional development for French cataloguers and metadata experts to help them make the transition.

In the Netherlands, the Dutch Network for Digital Heritage (NDE)[7] leads the bibliographic and cultural heritage data transition. It brings together the main players from the library, academic, media, museum, archive and architecture sectors with the aim to jointly implement the national strategy for digital heritage on behalf of the Ministry of Education, Culture and Science. The Dutch Ministry seeks alignment with the European Interoperability Framework (EIF),[8] which promotes interoperability of digital public services at EU levels. As a result, NDE developed the Digital Heritage Reference Architecture (DERA)[9] which, importantly, requires the application of linked data principles for the publication and linking of heritage data. The remit and ambition of the NDE are broader than just leading the bibliographic transition and address the full digital strategy for the heritage sector. On the other hand, its practical impact is limited to the voluntary adoption of best practices and guidelines, as it does not have a centrally coordinated approach or roadmap for the transition. However, the network is influential and successful in raising awareness, knowledge sharing, training professionals and deploying digital heritage coaches in its effort to mobilize the Dutch heritage community and its stakeholders.

The French and Dutch examples are very different in approach and scope, but both are efforts to mobilize the community towards a more unified approach to next generation metadata. They both showcase EU policies' impact on collaborative arrangements in the library and cultural heritage sector in Europe. The impact is obviously not the same in all EU member countries even if the guiding principle: "*digital-by-default, cross-border -by-default and open-by-default,*" is a shared one. The Italian round table session for example, mentioned Italian "Guidelines for Semantic Interoperability through Linked Open Data"[10] produced as early as 2013 and inspired by the EIF. However, Italian participants observed a lack of coordination between initiatives within the country and a fragmented cultural heritage landscape resulting in a variety of independent projects and a lack of common standards.

Improving the metadata in the supply chain

Round table participants reported how libraries are intentionally feeding external systems - e.g., the university's research portal, the ORCID-database, Wikidata, etc. - with their authority data to achieve critical mass. They are also systematically embedding persistent identifiers (PIDs) throughout the supply chain to rationalize it. This is particularly true for libraries and bibliographic agencies that act as registration agencies for identifiers, such as the International Standard Name Identifier (ISNI) for example, and who operate in the context of their national bibliography and legal deposit tasks.

More coordinated approaches that go beyond individual initiatives and institutional scale, are currently taking place in the UK to achieve efficiencies at scale. These were discussed during the English round table discussions. One such effort concerns the embedding of PIDs within the supply chain to get them out to where cataloguers can pick them up. To this end, the British Library is collaborating with publishers to encourage the adoption of ISNIs upstream. As a result, publishers are starting to use them in their ONIX feeds. Also, cataloguing agencies in the UK are being supplied with ISNIs so that they can embed them in the metadata at source, in the cataloguing-in-publication (CIP) metadata, that they supply to libraries in the UK. Efforts are also underway to work together with OCLC to systematically match ISNI entries against VIAF entries, and to provide a reconciliation file to enable OCLC to update the VIAF with the most recent ISNI. These identifiers could then be fed through to the Library of Congress, who can use these to update their LC/NACO Authority File. As one of the participants said:

"With 6 million files to update, this is a perfect example of a leading edge dynamic next generation metadata initiative that will have to overcome the considerable challenge of scalability for it to succeed at a global level."

In this context, the UK-HE-wide approach to funnel authority data in a coordinated way to the LC/NACO Authority File was also mentioned.

Finally, Plan-M[11] was briefly mentioned. It is an initiative, facilitated by Jisc, to rethink the way that metadata for academic and specialist libraries is created, sold, licensed, shared, and re-used in the UK. It is a conversation between libraries, suppliers and intermediary organisations to streamline the metadata marketplace in the UK so that it is more coherent, transparent, robust and sustainable.

Managing multiple scales

It was fascinating to hear about these country-initiatives on the one hand and to hear about the challenges faced by individual libraries striving to effectively move to the next level of metadata management, on the other. For the latter, often the most convenient and attractive way to overcome the issues they face, is to connect to national or global scaled efforts. Of the latter, two were most mentioned:

- LC/NACO Authority File, one of the largest international authority files maintained cooperatively, and which is massively being enriched with relevant identifiers.
- OCLC's Shared Entity Management Infrastructure, which is being built at global scale to provide a centralized infrastructure for managing shared library data about entities such as person names and bibliographic works.

These global components were also considered important to link up with for the country-initiatives. The challenge then becomes one of managing multiple scales and global interoperability. Andrew Pace, Executive Director, Technical Research at OCLC, described the landscape as a "*semantic continuum*" – portraying the challenge of managing multiple scales as one of "*bridging the effort between the short tail and the long tail*", in other words, between scaled effort and localized domain and collection expertise. To that end, libraries need to "*balance large, shared, homogenous collections and data, while accounting for a myriad of de-centralized and heterogenous collections*".

Conclusion

The round table conversations gave participants an interesting glimpse into the relief formation of the newly emerging landscape of the next generation of metadata. Unsurprisingly perhaps, but noteworthy nevertheless, national scale programmes - in France, the Netherlands and the UK - stood out as most intentional in mobilizing community, from the many different projects and initiatives discussed during the round tables. Participants in the Italian-, German- and Spanish-speaking sessions expressed a strong desire to move their next generation metadata efforts forward in a more concerted way. They clearly saw the benefits of cooperating more strongly at a national level and, also internationally, especially when cooperating with those further ahead. They expressed the wish to see OCLC help them organize more discussions like this at the country level and to play an active role in this moment of transition. This message did not fall on deaf ears, and at OCLC we are planning next steps with the community to keep the momentum going.

References

[1]OCLC Research Discussion Series on the Next Generation of Metadata (23 February – 13 April 2021): https://www.oclc.org/go/en/events/next-generation-of-metadata/Register.html

[2] Smith-Yoshimura, K. 2020. *Transitioning to the Next Generation of Metadata*. Dublin, OH: OCLC Research. https://doi.org/10.25333/rqgd-b343

[3] The OCLC Research Library Partnership (RLP) Metadata Managers Focus Group homepage https://www.oclc.org/research/areas/data-science/metadata-managers.html

[4] OCLC Research. Equity, Diversity, and Inclusion Initiatives (viewed 7 June 2021). https://www.oclc.org/research/areas/community-catalysts/edi.html

[5] CILIP Conference 2020, session 'Transitioning to the Next Generation of Metadata' (19 November 2020): https://pheedloop.com/cilipconf20/site/sessions/?event=cilipconf20&section=38349&id=HxkkpJ

[6] The Bibliographic Transition in France webpage (in English): https://www.transition-bibliographique.fr/enjeux/bibliographic-transition-in-france/

[7] The Dutch Network for Digital heritage homepage: https://netwerkdigitaalerfgoed.nl/

[8] Annex 2 to the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'European Interoperability Framework – Implementation Strategy'. COM(2017). English version. http://eur-lex.europa.eu/resource.html?uri=cellar%3A2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_3&format=PDF

[9] Digitaal Erfgoed Referentie Architectuur webpage (in Dutch): https://netwerkdigitaalerfgoed.nl/activiteiten/dera/

The Need to Develop a Collection of Subject Terms on Nepal's Diversity

Ms Bina Vaidya, Former Associate Librarian, Tribhuvan University Central Library

We have a large number of subject headings reflecting Nepal's diversity, including language diversity, diversities in festivals, musical instruments, ethnic groups, cultures, dresses, and so on. Although these exist and are in use in Nepal, people outside Nepal do not have access to them. That is why those headings are not found in the collection of Library of Congress Subject Heading (LCSH) list. In the same way it is not available in Dewey Decimal Classification (DDC) scheme. When there is no subject heading reflecting a topic to align to, DDC editors also cannot include DDC class numbers references to in their published editions. However, Nepalese professional librarians can collect these headings and send them to Library of Congress office to include in the LCSH list. In the same way we can also request DDC editors to include references to class numbers in the official edition. It is not necessary that all compiled headings have to be included in the subject heading list and also in DDC in the same time. But it seems high time to develop the collection of vocabularies on Nepal's diversity and to have them available on these central global resources.

My experience while working in the Tribhuwan University Central Library (TUCL)

I have worked in TUCL for 25 years. I have worked there from when I entered the profession in 1977 until I retired. As I was interested more in the technical work in the technical section of the library from the very beginning I joined this library to work in that area. I used to work in the technical section and became the its chief after the death of librarian Purna Prasad Amatya. I got an opportunity to learn many technical works from him when I was working under him.

Although this library is a university library, it also functions as a national library and a public library providing service to all the people of Nepal. This library is the largest, well managed and modernized library in Nepal. The library holds a large number of collections of books, journals, rare books and manuscripts and personal collections of historians, politicians, teachers, educationists, etc. Among these the most important and largest one is the "Singha Collection", donated by Singha Sumsher with beautiful British furniture. The "Nepal collection" is the next most important special collection of the library. More than 50,000 documents on Nepal including Nepal's diversity is available in this collection. This is the only library in Nepal where such a large collection of Nepal is available in Nepal.

Large numbers of books and library materials on various subjects have to be processed by the technical section and we use LCSH for assigning subject headings. And while using LCSH we are facing problems sometimes as not all the subjects we need have entries in the scheme, especially on subjects relating specifically to Nepal. This is not as great an issue in small libraries with correspondingly small collections.

The Library receives the library materials from purchase and gifts, with 60% of materials received from national and international donors, book publishers, etc.

One of the reasons of the increasing size of the collection in Nepal is because of this library being a National Agency for distribution of ISBN numbers. Publishers have to submit one copy of each publication to the library. To handle such a large collection requires the use of many subject headings, but the headings given in the LCSH are not enough for TUCL specially on Nepal and Nepal's diversity. As a result the TUCL has to maintain its own Subject Authority File in which all headings that have been given to materials are added. Besides the headings themselves, many related decisions have to be recorded as scope notes in the technical section. For that we have maintained a Technical Section Manual.

Maintenance of Library Manual and Subject Authority File within the Technical Section

We record whatever decisions and policies are made while processing the library materials in a manual. The staff of the technical section have to follow the instructions given in the manual to bring uniformity and

consistency in the library work. Although large numbers of headings are available in LCSH we do not find the headings we need. What we need especially are whole collections of headings on Nepal's diversity and topics specific to Nepal related to language, culture, ethnic groups, festivals, dresses, etc. However, because those subjects are unique to Nepal they are not known outside Nepal.

As far as possible the cataloguers attempt to determine terms that would be used by users to search and discover the concepts, subjects and all pertinent materials they are interested in getting finding.

As a result it has been decided to maintain a Nepalese Subject Authority File in which we have recorded all the headings created for assigning those terms.

The provision of Dewey Decimal Classification DDC numbers for the provinces of Nepal by expanding number for Nepal 954.96

DDC has assigned DDC number 954.96 for Nepal and this has remained the same throughout its publication history whereas the DDC for India 954 has been expanded significantly and includes separate numbers for some major cities of India such as Calcutta, New Delhi, Bombay, Bangalore, etc. According to the Federal Democratic Republic of Nepal 2015, Nepal has been divided into 7 provinces. On the basis of this expansion of India's number 954, similarly Nepal's number 954.96 should also be expanded, at least to the extent of assigning numbers for each of these provinces. Fortunately, I had got an opportunity to discuss this topic during a meeting with an OCLC editor while attending IFLA OCLC meeting in Malaysia in 2018. She was convinced and agreed to and accepted my proposal to provide separate numbers for provinces of Nepal expanding Nepal's number 954.96 (954.961 for Province no. 1, 954.962 for Province no. 2, 954.963 for Province no. 3, 954.964 for Province no. 4, etc.)

Before that also we had prepared a subject structure for Nepal with the following hierarchy:

Nepal – Development Region – Anchal (zones) – Districts – Villages – Towns and villages

Starting from 954.96 for Nepal 954.961 to the development region, 954.9611 to the Anchal or zone, 954.96111 to districts. For villages and towns, the name of the village with first alphabet will include after district number.

This guide line has been prepared according to the political structure of that time. We sent it to DDC system office OCLC Online Computer Library Center, Dublin. Although there was no reply received these further expansions were used as local expansions to create individual classifications for Nepalese towns and villages.

The Need to Develop Vocabularies on Nepal's Diversity

As discussed earlier there was an identified need to develop and collect the headings on Nepal's diversity to keep them all in one place. We discuss here the diversities of linguistics, culture, festivals, ethnic groups, dress.

Linguistic Diversity in Nepal

Our linguistic diversity is closely related to ecological and cultural diversity. There are four main language families being using in Nepal. Like the mountains, these were formed by the collision between the Indian and Tibetan continental plates, the language, religion, ethnicity and culture of the Himalayan Kingdom have also been born of a collision between North and South. Those four language families are Indo-Aryan, Tibeto-Burman, Munda and Dravid. According to National census, there are 12 major languages related to those four language families. In the last five decades of modernization more localized languages are in decline. The official language is Nepali and other major languages in use are:

- Nepali
- Tharu
- Magar

- Hindi
- English
- Tamang
- Maithili
- Bajika
- Newar
- Bhojpuri
- Urdu
- Sanskrit

In the last five decades of modernization, Nepal has seen the decay of many local language and dialects. The names of many of Nepal's languages were derived from place of origin and have their own historical traditions.

Nepalese Cultural Diversity

Nepal is one of the most famous nations for its diversity in cultures, geographical settings and many other essential aspects throughout the world. In Nepal more than 100 ethnic and caste groups exist and have their own cultures.

Nepal is a small country but broad in its diversity. The way of life of the people's life, food, clothing and even occupation are guided by the culture of Nepal. The culture of the people includes the code of manners, rituals, clothing of the people (dress), the spoken language, the system of beliefs, norms of behavior, etc. Nepal has diverse cultural practices, numerous rituals, beliefs, social values, festivals, arts, architectures, rites and other aspects and customs. Cultures of Nepal are different from one part of Nepal to another and one cast to another.

Festivals in Nepal

Because of Nepal's diversity in culture the festivals in Nepal are diverse too. Numerous festivals are celebrated every year. Many people in Nepal celebrate festivals and jatras almost every day. These festivals and jatras have religious or traditional values attached to them. Some of the significant festivals of Nepal are listed below:

- Dashain
- Tihar
- Bisket Jatra
- Chhath
- Lhosar
- Holi
- Mahashivaratri
- Buddhajayanti
- Teej
- Janai Purnima

Among them the festival with the longest history, held to be most auspicious of all those held in Nepal is Dashain. The celebration of each festival in Nepal is equally enjoyable and exciting.

Nepal's Music Instruments

Nepalese musical instruments are related to Nepalese culture and religion and are as diverse as the various ethnic groups of the country. Those musical instruments have many unique tunings and rhythms of their own to share with the rest of the world. These instruments have great importance in Nepalese cultural and society and are part of the cultural heritage of Nepal.

Some of the popular musical instruments of Nepal are as follows:

- Madal
- Khainjadi
- Bansuri
- Murali
- Sarangi
- Panchebaaja
- Pungi
- Dhyangro
- Hudku
- Dholakh
- Tunaga
- Yalambar
- Ektara
- Musak

These musical instruments are used along with singing of songs during festivals ceremonies, jatras and other cultural programmes. Newar community mostly enjoy and have musical cultural tradition.

Ethnic Diversity

The History of Nepal traces the history of migration of large numbers of people of Asian ethnicity from Tibet, Indo-Aryan ethnicity from Northern India adding to descendants from the original settlement of Nepal. The migrated Asian group and Indo-Aryan group have introduced a diverse linguistic ethnic, and religious combination to Nepal.

The earliest recorded rulers were of Indo-Aryan and Hindu background. Tibeto Nepali groups include Tamang, Rai, Limbu, Bhotia, Sherpa and Sunwar and Indo-Aryan groups are Brahmin, Chhetri. Third ethnic groups include Newars and Tharus and they are believed to have settled in Nepal before the Tibetan and Indo-Aryan migration.

Nepal has 126 castes and ethnic groups speaking and as many as 123 languages including the languages derived from the place they live. In Nepal there are about 101 ethnic groups or speaking over 92 languages. These are some of the following ethnic groups:

- Chhetri
- Brahmin
- Magar
- Tamang
- Newar
- Kami
- Muslim
- Rai
- Yadav
- Gurung
- Thakuri
- Damai
- Limbu
- Sarki
- Teli
- Chamar
- Harijan

- Koiri
- Kushwaha and others.

This is according to the report of National Census of 2011.

Diversity in Dress

Like all countries there are plenty of traditional costumes in Nepal. Those traditional costumes are based on caste and ethnicity. However, Daura Suruwal of men and Guniyocholo of women are the national costumes of Nepal and are worn by Nepalese of all sections, class and ethnicity. Traditional costumes are also symbols of the art and culture of Nepal.

Nepali Traditional Dress for Men:

- Daura Suruwal
- Dhaka Topi
- Dhoti Lungi

Nepali Traditional Dress for Women:

- Guniyocholo
- Kurta Suruwal
- Blouse
- Sari

And various other ethnic costumes.

Conclusion

In brief I have described Nepal's diversity on linguistics, culture, festivals, ethnic groups, musical instruments and dresses. While cataloguing those library materials there is a need to assign whole groups of accurate subject headings and but these are not available in the Library of Congress Subject Heading (LCSH) index, which creates problems. The solution has been cataloguing these library materials using a maintained national Subject Authority File. This Authority file is not currently available outside of Nepal. My intention is to send those collections of headings on Nepal's diversity to be included in the LCSH in the coming edition through applying to the office of the Library of Congress, Washington.

Similarly, I will also attempt to integrate local classifications to include provincial sub divisions of Nepal's DDC number 954.96 as mentioned earlier. Again, this will be done by contacting the editor of DDC. However, before these integrations, my task will be to collect in detail all collections of Nepalese diversity related subject headings.

BIC's Metadata Capabilities Directory

Peter Mathews, BIC Consultant, lead for the project

Book Industry Communication is shortly to launch its Metadata Capabilities Directory (MCD) after testing and trialling with a small group of 'Pioneer' members. The new facility, which is part of BIC's 'Metadata Map' project, has been under development for the last couple of years, but it is now hoped to be rolled-out to the entire BIC membership within the next few months.

So what is the MCD and why has it been created?

The project originally started as a result of repeated questions raised at BIC's Metadata Committee. Whilst it was recognised that there were plenty of excellent 'Best Practice' guidelines available for the creation and supply of product data, it was much less clear that we knew what happened to metadata after that, once it had entered the metadata supply chain.

Do we really know what is actually sent by individual suppliers and how? Are we able to be certain what data recipients do with the data when they receive it and apply it to their systems? Or if they send it back out again, what changes are made to format or content? Perhaps there are useful data attributes that could be sent that are not? Or attributes that *are* sent, but are ignored by receivers?

The conclusion was that if we could establish a way to allow organisations to share what they actually do with the metadata they create, receive or pass-on we might learn some practical answers to these questions and improve the data supply chain for everyone.

The original objective was to "*Build a sustainable Metadata Map of the UK metadata supply chain and to share the aggregated results with all participants to provoke positive change in data behaviours by users."* (LBF launch in 2018)

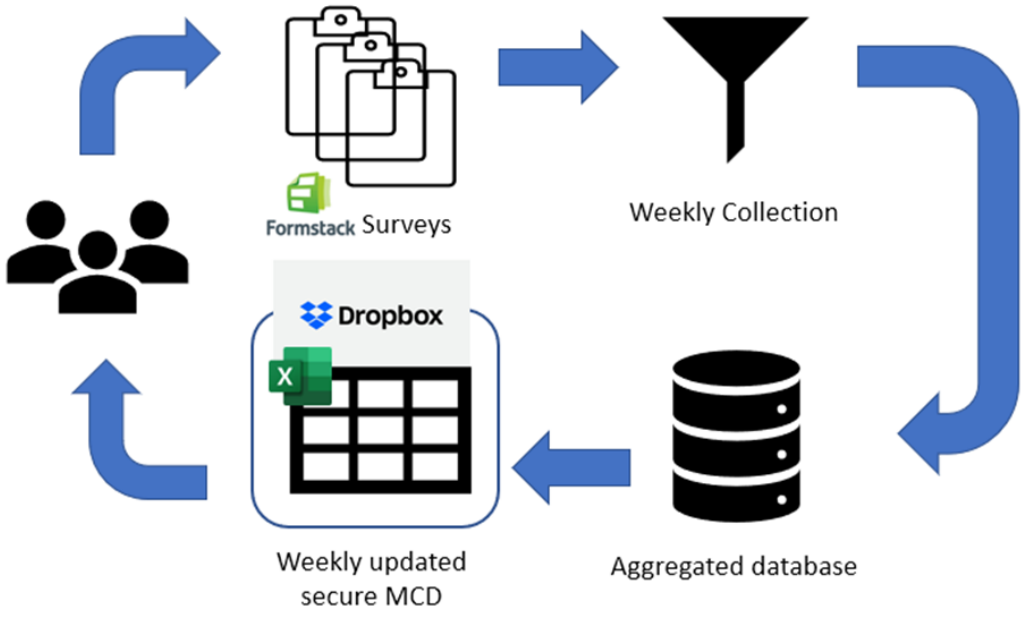
However, we were unsure how much information organisations would be willing to provide about their data activities and how detailed we could legitimately and practically make our questions. How could we best present the resulting information to make it useful to the whole supply chain by providing top-level overview analysis of the data supply chain at the same time as the drill-down detail of organisation-to-organisation and specific data attributes? Perhaps most importantly, did this project have sufficient widespread support across the supply chain to generate useful data? How could we unify the very divergent views on what this project should deliver?

Our ambition therefore needed to be tempered with a dose of realism. We have created a detailed survey for participants to answer which we believe is sufficiently detailed, but practical for all organisation types active in the data supply chain to be able to answer. It is important to note that we are not looking to capture the actual metadata being shared within the supply chain, but rather we are seeking to understand the policies and system limitations that define and shape metadata behaviours within our industry.

We also recognised that we could not justify spending too much of BIC's members money on developing a very sophisticated solution until we better understood what was wanted and what was possible. To find this out we needed to create sufficient trust and demonstrate value amongst potential participants to enable them to share information in the first place.

The MCD system

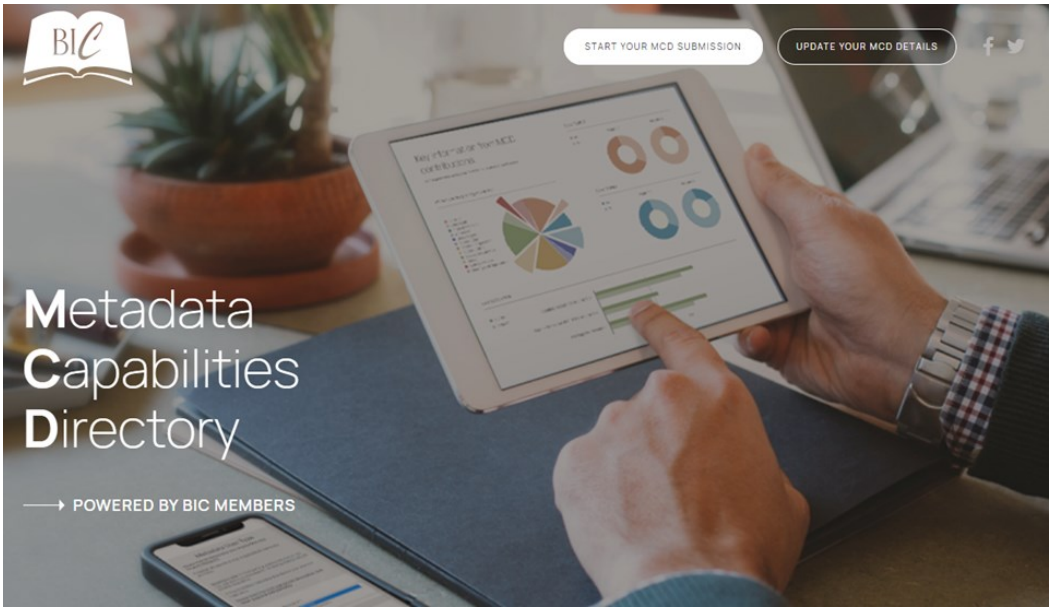
The result is the MCD, a simply-presented summary listing of responses to our survey, detailing the data activities of each of the participants, that is password-protected and only accessible to BIC members who have contributed to the directory.



Participants complete a Formstack survey, covering their full range of metadata activities, whether they are a data supplier, recipient, or both. Newly submitted surveys are downloaded weekly and added into a secure database, from which we extract a simple spreadsheet of all submissions. We make this available securely within Dropbox, so that only contributors can view it.

Organisations needing to update their submission do so by the same process, using an update survey which allows them to change only the information they need to change, which is then collected, aggregated, and re-presented to users in the updated MCD each week.

To attract new participants to the MCD, we will shortly launch a portal webpage within the BIC website which provides an introduction to the MCD, hosts the documentation created for users, provides the links to the Formstack surveys and showcase a few key overview statistics collected from MCD submissions.



How does the MCD work?

The scope and complexity of the product metadata that is used and communicated within the supply chain is significant and it follows that we needed our survey to reflect this to make the MCD useful. Our question set has therefore been designed, and refined by our Pioneer Group testers, to allow a comprehensive overview of the whole range of product data created and moved within the supply chain and to enable true insight into the processes that shape and maintain it. The scope is the same for both Data Suppliers and Receivers, with questions and responses reflecting each perspective. Most questions offer a set of standardised multiple-choice options. This helped to frame the type of responses we were looking for, but also provided a more practical way of aggregating the responses in the MCD. Filtering on standard terms is much easier than trying to make sense of free-text.

Nevertheless, there is still plenty of opportunity for organisations to elaborate or refine their responses with free-text, supplementary boxes available in most sections of the survey. We hope that the MCD becomes a trusted source of comprehensive information for all contributors to use and we believe that this is best achieved by encouraging a spirit of honesty and full disclosure.

We have taken considerable steps to ensure the security of all data provided, which we believe is paramount to the success of the MCD. We have required each contributing organisation to nominate a single authorised individual who takes responsibility for that organisation's submission and is the authorised user to access the MCD file on Dropbox. The spreadsheet file is read-only to emphasise that data supplied into the MCD should only be seen and used within the context of the MCD by all users and to ensure that only contributors to the MCD are able to see its contents. In addition, each participating organisation must create a Unique ID which must be used to ensure any future updates are aggregated with their previous responses. All new submissions are validated by BIC with a completeness threshold required to met before access is granted to the MCD spreadsheet on Dropbox.

Why only a spreadsheet? It helps keep the cost down – we have not needed to pay for the design or development of an interface to deliver the contents – and everyone knows how to use standard spreadsheet functionality like 'Find' and 'Filter' to enable the data to be easily interrogated.

How will MCD users benefit?

We recognise that organisations are unlikely to participate just on the promise of improving the whole supply chain. They will want to know what they can get out of it themselves to benefit their own business and the effectiveness of their own metadata.

Pioneer Group contributors have already identified that simply by taking the survey they have shone a light into areas of metadata use in their own business that warranted attention. We believe however that the main benefit will be by MCD users benchmarking their own activities against other similar organisations, competitors and/or data partners and against the overall trends of all participating organisations, and thereby able to identify beneficial improvements to their own systems and practices. In addition, it may be possible to justify the cost of system development and prioritise the most valuable improvements to that organisation. Nevertheless, in time, we also believe that there may be valuable lessons to be learnt when looking at overall data and trends in the future that could highlight data usage improvements we can all make within the supply chain.

Launch of the MCD

We have already started receiving live, full submissions from our Pioneer Group and more are promised in the very near future. The system is successfully aggregating these and creating the system reports including the MCD itself, and this will shortly be shared with those contributors for the first time. The dedicated MCD webpage on the BIC website is also being updated routinely behind the scenes and will be published when we formally launch the MCD soon. We look forward to inviting BIC members to participate soon, and enabling the expected, shared, benefits of the MCD to become a reality for all our members and the wider supply chain.

Wikidata: What? Why? How?

Helen Williams, Metadata Manager, LSE Library

During 2019 I noticed that Wikidata was a growing topic of conversation in the world of metadata. Its power to create links and show relationships between entities stirred my interest and I was keen to investigate potential benefits to the Library and the wider institution. Wikidata takes the collaborative creation and management of metadata beyond the Library to a global landscape, and this issue of C&I, focused on collaboration, seemed to offer a timely opportunity to write about our Wikidata progress, even though we are still learning.

What is Wikidata?

Wikidata is a structured database operating as the central data store for all Wikimedia projects.[1] It is a 'free and open knowledge base that can be read and edited by humans and machines'[2] and is multilingual, supporting global access to information. Google Knowledge Graphs, digital assistants such as Alexa and Siri, and Wikipedia Infoboxes are all populated, in part, with information harvested from Wikidata so its content has a real impact on discovery.

Why is the Library interested in it?

Sharing library metadata with Wikidata means that unique identifiers can be minted for library content and the entities within it, pulling it into the Linked Open Data ecosystem and connecting it with unique identifiers from external knowledge systems. This starts to create bridges between currently siloed domains, which in turn impacts search engine results, making the content more widely accessible and enabling new connections and discoveries which can support global research.

How did I begin working with Wikidata?

Working with Wikidata has involved a steep, but very rewarding, learning curve. Some institutions are able to draw on the expertise of a Wikimedian in Residence, which is not our situation, so I very much started from scratch, reading articles and webpages and watching presentations and videos online. Content produced by Jason Evans,[3] Martin Poulter,[4, 5] Nick Sheppard,[6] The University of Edinburgh and Wikimedia UK,[7] and Harrison Pim[8] was particularly useful, as were discussions within the OCLC Research Library Partnership Metadata managers group, some of which are summarised on the Hanging Together blog.[9] To familiarise myself with practical work on Wikidata I started with Dan Scott's blog post about creating and editing libraries in Wikidata[10] and edited the existing item for the British Library of Political and Economic Science,[11] adding some new properties and identifiers. I also created some new items, such as LSE Digital Library.[12] which I could link using the 'has part'[13] and 'part of'[14] properties. I next looked at mapping different content types to Wikidata and created data models for LSE Digital Library collections, blogs, OA journals and online exhibitions, as well as for corporate bodies related to the Archives. I also created some corresponding Wikidata items for each model.[15, 16, 17, 18, 19] This provided some ideas for potential avenues of work, which I summarised by area of focus, including organisational, community, research, theses (inspired by Martin Poulter's work at Oxford University[20]), open access, digital and archival. A few interested colleagues kindly reviewed and discussed these options with me.

Having learnt the basics, I summarised for non-metadata colleagues how Wikidata could be of value to the Library, and why it was worth the investment of our time. Alongside this I proposed that the theses data in LSE Theses Online (LSETO) would be a valuable dataset on which to begin experimental work. This is a boundaried dataset of about 4000 records, the incorporation of which on to Wikidata would offer value to early career researchers and alumni by promoting their work, and would allow us to expose DOIs in Wikidata as these were added to the repository. Our theses deposit agreement allows us to share the metadata[21] and this project would not expose any data which had not already gone through internal processes for being made available in the public domain. As Wikidata is outside the control of the Library there are potential concerns about our data being edited by external users. However as we are using Wikidata to extend the reach of metadata that is already held in an internal library system we are not exposing ourselves to any risk of data loss; and if edits in Wikidata created additional links with related entities and sources this would be positive. Trusting that the vast majority of the Wikimedia community will be acting in good faith, and as our metadata is not controversial, I am not expecting problematic edits. We will, of course, need to keep this under review, address any issues arising and note any lessons learned so that we can take them into account for any future projects.

The work was approved and having contacted the Wikimedia Foundation to discuss our planned data donation,[22] I embarked on teaching myself how to export and manipulate data for bulk import into Wikidata, how to reconcile LSE names with existing Wikidata identifiers and how to link LSE data to identifiers in external datasets. The process I have developed is outlined below.

Exporting data from LSE Theses Online

For the purposes of learning and experimenting I wanted to work with relatively small sets of data, so I have been exporting data from LSETO one year at a time, by multiline CSV, deleting unnecessary data and concatenating various columns where required.

Preparing LSETO data for import to Wikidata

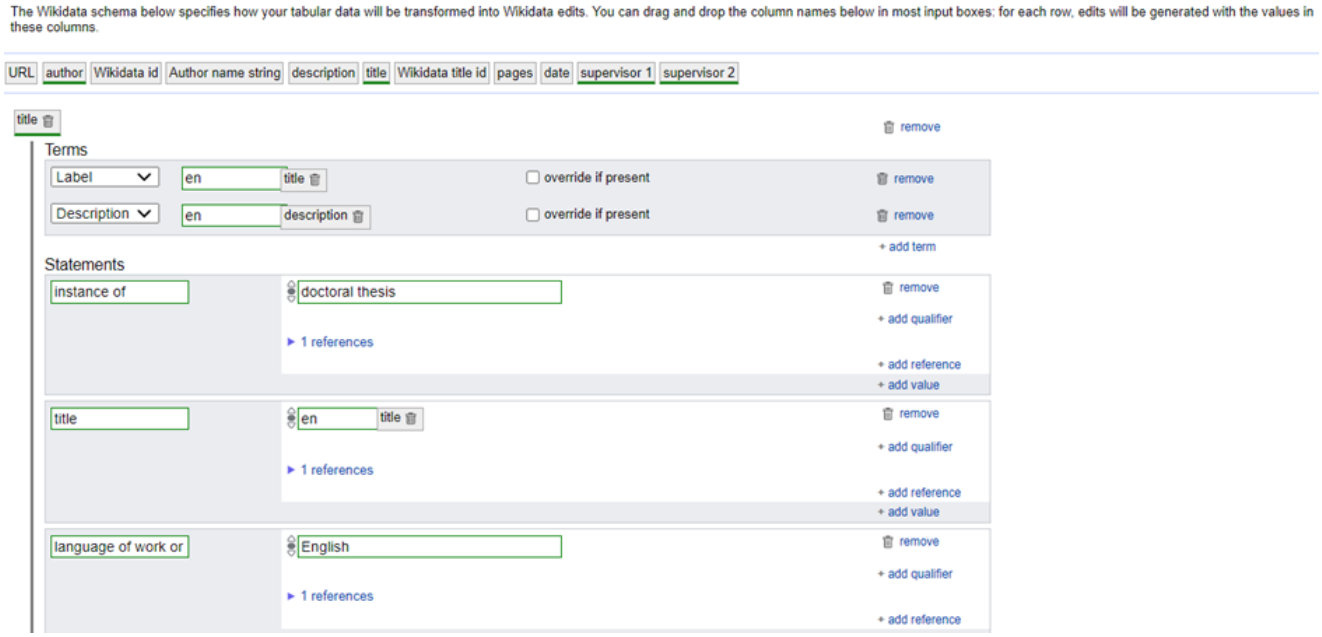
I next wanted to format the data for upload to Wikidata using QuickStatements. I followed a YouTube tutorial from Sara Thomas[23] to learn about this, and I am very grateful for subsequent guidance from Simon Cobb who helped me to identify small formatting mistakes I was making which were preventing data upload. The table below represents our data model.

A	B	C	D	E	F	G	H	I
CREATE								
LAST	Len	"Title"						
LAST	Den	"doctoral thesis by author"						
LAST	P31	Q187685			S854	"url"	S813	+2021-mm-ddT00:00
LAST	P1476	en:"Title"			S854	"url"	S813	+2021-mm-ddT00:00
LAST	P2093/P50	"Author"/Qid	P1932	"Author"	S854	"url"	S813	+2021-mm-ddT00:00
LAST	P4101	Q174570	P184	supervi-sor Qid	S854	"url"	S813	+2021-mm-ddT00:00
LAST	P407	Q1860			S854	"url"	S813	+2021-mm-ddT00:00
LAST	P577	+YEAR-00-00T00:00:00Z/9			S854	"url"	S813	+2021-mm-ddT00:00:00Z/11
LAST	P1104	pages			S854	"url"	S813	+2021-mm-ddT00:00
LAST	P6216	Q50423863			S854	"url"	S813	+2021-mm-ddT00:00
LAST	P953	"url"			S854	"url"	S813	+2021-mm-ddT00:00
LAST	P6954	Q232932			S854	"url"	S813	+2021-mm-ddT00:00

- Each new record is identified by the presence of CREATE in column A.
- Each new statement is identified by the presence of LAST in column A.
- Column B identifies the statements to be created.
- Column C provides the data which those statements will contain.
- Len means Label in English, for which we are using thesis title.
- Text strings need to be surrounded with double quotation marks.
- Den means Description in English.
- The following properties are used to populate the Wikidata item for each of our theses:
    - P31 (instance of) Q187685 (doctoral thesis)
    - P1476 (title) – text string of thesis title, prefaced by en: to represent language
    - P2093 (author name string)
        * If the author is already represented in Wikidata P2093 is replaced with P50 (Wikidata item for author) and their Qid
        * If the name is formatted differently in LSETO the qualifier P1932 (stated as) is added in column D and the text string in column E
    - P1401 (dissertation submitted to) Q174570 (London School of Economics and Political Science)
        * If the doctoral supervisor for the thesis has a Qid the qualifier P184 (doctoral supervisor) is added in column E and the Qid in column F
    - P407 (language) Q1860 (English)
    - P577 (date) +YEAR-mm-ddT00:00:00Z/9 (for syntax see Wikidata[24])
    - P1104 (pages) – number of pages
    - P6216 (copyright status) Q50423863 (copyrighted)
    - P953 (full work available at URL) – LSETO URL
    - P6954 (access status) Q232932 (open access)
    - P365 (DOI) – not included in table above but added if already in LSETO
- S854, in column F, represents the reference URL required for each statement and column G is populated with the LSETO URL to generate that reference.
- Each reference URL requires an S813 (date retrieved) statement.

Once the metadata is formatted, as above, multiple records can be pasted into QuickStatements for upload to Wikidata.

I subsequently learnt that data can also be uploaded to Wikidata directly from tabular data in OpenRefine[25] and after some trial and error have been able to establish an upload process using this method too. This has required using the custom facet and transform functions in OpenRefine to manipulate the data into the appropriate format, including separating out author names with Qids requiring a P50 statement, and those requiring a P2093 author name string statement. I was then able to create a Wikidata schema in OpenRefine to transform the tabular data into Wikidata statements. The following 2 images show the initial canvas from which the schema can be created, and then a subsection of the schema I have created.



This schema can be exported from OpenRefine in JSON and then imported again for use on subsequent theses datasets in OpenRefine. While I have been working on this one of my colleagues is working on efficiencies in the processing and manipulation of LSETO data into our QuickStatements model. Having refined both processes we will then be able to assess which is the most effective method for upload.

Data reconciliation

OpenRefine has been a vital part of the project for two aspects of data reconciliation and I am indebted to Owen Stephens for his YouTube tutorials on this.[26] We are using P2093 (author name string) to enter author name (for example, https://www.wikidata.org/wiki/Q106600336), but this does not provide any additional author details or further discovery paths. If, however, the author is already represented in Wikidata we can use P50 (Wikidata item for author) which creates a hyperlink, allowing the discovery journey to continue more easily. Taking https://www.wikidata.org/wiki/Q105822883 as an example it can be seen that the use of P50 facilitates the discovery of additional information about the author, where this already exists in Wikidata, including various identifiers in other datasets. There are a number of authors and supervisors in our data who are not already represented in Wikidata and, although it may be valuable to create Wikidata items to represent them, limited time and resources means our initial focus is on exposing the bibliographic data. The creation of new items for living persons would also require additional ethical considerations[27].

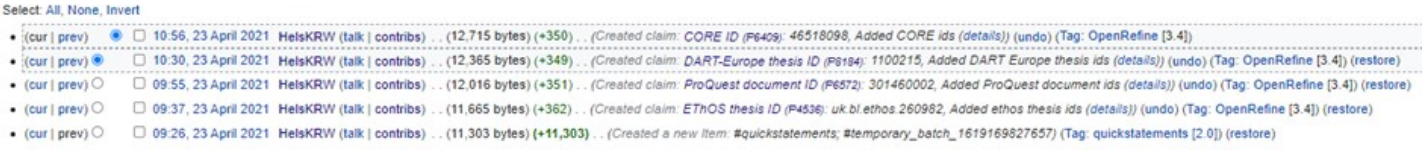
In instances where both author and supervisor have a Qid this data can be 'round tripped'. For example, from the item on the left, clicking on 'Maurice Bloch' links to his Wikidata item where P185 (doctoral student) has been added with the identifier for 'David Lan'. On following his hyperlink users discover P1026 (academic thesis) with a hyperlink back to the item for the thesis title. For each thesis containing P50 one of my colleagues adds P1026 (academic thesis) to the author's Qid page and checks for the existence of/adds P69 (educated at) and, if both author and supervisor have Qids, links these via addition of P184 (doctoral advisor) and P185 (doctoral student).

It would be very time consuming to search Wikidata manually for each name to see if it is already represented, but the OpenRefine reconciliation function can automatically check a spreadsheet of names against Wikidata and return the relevant Qids. Some matches are made automatically (though should be checked for false positives), while others have potential matches for investigation.

The second aspect of data reconciliation which OpenRefine has facilitated is linking the newly added LSE theses content to identifiers in external datasets. I have been able to obtain thesis titles and identifiers to support automating the process for the following datasets:

Dataset	Wikidata Property	Process
EThOS	P4536	Extract LSE data from Ethos dataset[28] Edit Ethos URL to match Wikidata formatter URL[29]
DART-Europe	P8184	Contacted UCL Library Services for LSE dataset containing theses titles and corresponding identifiers
ProQuest	P6572	Extract LSE data via advanced search on ProQuest Dissertations and Theses[30]
CORE	P6409	Extract LSE data via CORE API[31] Convert JSON results to CSV

Once the titles and corresponding identifiers are in a spreadsheet a project can be created for each dataset in OpenRefine. The titles can then be reconciled with Wikidata and the facets filter used to select matched titles before creating a Wikidata schema in OpenRefine which uses label, language and title to identify the correct Wikidata item, adds statements containing the corresponding identifiers and uploads edits directly to Wikidata. The history tab of these Wikidata items shows that the original upload was from QuickStatements, and the identifiers added via OpenRefine.[32]



Following the example of Martin Poulter's SPARQL query to show Oxford theses in Wikidata[33] I created a similar query for LSE.[34] Where the author has a Wikipedia page we add the thesis title to the main body of text, with an external link to LSETO, supporting this with an inline citation using the {{cite thesis}} template, and adding the thesis to an existing infobox where applicable.[35]

At the time of writing I have added 589 theses to Wikidata, so there is still much to do, not only in terms of adding further content and refining the process, but also to consider how we might visualise the data and to investigate what kind of impact the work has had on thesis downloads from the repository. It will be evident that I am still learning and that the work is still in progress, but I see this as an exciting piece of work for the Library, with the potential to make our content more widely available and to provide an opportunity for Library staff to learn new skills. I hope that this project will go on to provide a helpful foundation from which to explore the value of Wikimedia engagement both to the Library and to LSE.

Thanks to my Library colleagues Andy Jack, Helen Porter and Neil Stewart who have been valuable 'sounding boards' for this work, and to my Metadata colleagues Ryan Kermode and Gemma Read who have been willing to get involved with this experimental area of work for our team.

References

[1]Wikidata (2019), *Welcome to Wikidata,* www.wikidata.org (Accessed 11/5/2021)
[2]Wikidata (2019), *Welcome to Wikidata,* www.wikidata.org (Accessed 11/5/2021)
[3]Evans, Jason (2020), *Creating and enriching Linked Data with Wikidata,* CILIP Metadata & Discovery Group conference 2020, https://www.youtube.com/watch?v=RApW2x4KJfw (Accessed 19/10/2020)
[4]Poulter, Martin (2019), *Wikidata: GLAM/Oxford,* https://www.wikidata.org/wiki/Wikidata:GLAM/Oxford (Accessed 19/10/2020)
[5]Poulter, Martin and Nick Sheppard (2020), *Wikimedia and universities: contributing to the global commons in the Age of Disinformation.* Insights 33 (1): 14, http://doi.org/10.1629/uksg.509 (Accessed 19/10/2020)
[6]Sheppard, Nick (2019), *Wikimedia in universities.* Leeds University Library blog, 1 Feb 2019, https://leedsunilibrary.wordpress.com/2019/02/01/wikimedia-in-universities/ (Accessed 30/4/2021)
[7]Wikimedia UK and the University of Edinburgh (2020), *Wikimedia in education,* https://open.ed.ac.uk/wikimedia-in-education/ (accessed 19/10/2020)

[7]Wikimedia UK and the University of Edinburgh (2020), *Wikimedia in education,* https://open.ed.ac.uk/wikimedia-in-education/ (accessed 19/10/2020)
[8]Pim, Harrison (2020), *Concepts, Wikidata etc.* Wikidata and Cultural Heritage Collections, Science Museum Heritage Connector webinar, 19 Jun 2020,https://www.youtube.com/watch?v=kZGcnL8gI6Y (Accessed 19/10/2020)
[9]OCLC (2020), *Hanging Together: Wikimedia,* https://hangingtogether.org/?cat=202 (Accessed 30/4/2021)
[10]Dan Scott (2018), *Creating and editing libraries in Wikidata,* https://coffeecode.net/creating-and-editing-libraries-in-wikidata.html (Accessed 11/5/2021)
[11]Wikidata (2021), *British Library of Political and Economic Science,* https://www.wikidata.org/wiki/Q2371017 (Accessed 11/5/2021)
[12]Wikidata (2021), *LSE Digital Library,* https://www.wikidata.org/wiki/Q96354844 (Accessed 11/5/2021)
[13]Wikidata (2021), *has part,* https://www.wikidata.org/wiki/Property:P527 (Accessed 11/5/2021)
[14]Wikidata (2021), *part of,* https://www.wikidata.org/wiki/Property:P361 (Accessed 11/5/2021)
[15]Wikidata (2021), *Oral evidence on the suffragette and suffragist movements,* https://www.wikidata.org/wiki/Q100380678 (Accessed 12/5/2021)
[16]Wikidata (2021), *LSE Impact Blog*, https://www.wikidata.org/wiki/Q77710848 (Accessed 12/5/2021)
[17]Wikidata (2021), *Journal of Illicit Economies and Development*, https://www.wikidata.org/wiki/Q96715673 (Accessed 12/5/2021)
[18]Wikidata (2021), *Making Modern Women*, https://www.wikidata.org/wiki/Q105556001 (Accessed 12/5/2021)
[19]Wikidata (2021), *Society for the Ministry of Women in the Church*, https://www.wikidata.org/wiki/Q82749481 (Accessed 12/5/2021)
[20]Poulter, Martin (2017), *A step forward in the sharing of open data about theses.* Bodleian Digital Library, 19 Jul 2017, http://blogs.bodleian.ox.ac.uk/digital/2017/07/19/a-step-forward-in-the-sharing-of-open-data-about-theses/ (Accessed 11/5/2021)
[21]LSE (2021), *LSE Theses Online-FAQs,* https://etheses.lse.ac.uk/faq.html (Accessed 11/5/2021)
[22]Wikidata (2021) *Wikidata: data donation,* https://www.wikidata.org/wiki/Wikidata:Data_donation (Accessed 11/5/2021)
[23]Thomas, Sara (2020), *How to use QuickStatements - a tool to bulk upload data onto Wikidata,* https://www.youtube.com/watch?v=QI7gC91eWss (Accessed 11/5/2021)
[24]Wikidata (2021) *Help: QuickStatements* https://www.wikidata.org/wiki/Help:QuickStatements (Accessed 11/5/2021)
[25]OpenRefine (2021), *Wikidata,* https://docs.openrefine.org/manual/wikidata (Accessed 11/5/2021)
[26]Stephens, Owen (2017), *Reconcilliation [sic] in OpenRefine ,* https://www.youtube.com/watch?v=q8ffvdeyuNQ&list=PL_0jeq3PjvtADzbovAgHNzOFvOlyF6uL1(Accessed 11/5/2021)
[27]Wikidata (2021), *Wikidata: Living people, https://www.wikidata.org/wiki/Wikidata:Living_people* (Accessed 11/5/2021)
[28]EThOS (2020), *EThOS* dataset, https://doi.org/10.23636/1188 (Accessed 06/11/2020)
[29]Wikidata (2021), EThOS thesis ID, https://www.wikidata.org/wiki/Property:P4536 (Accessed 11/5/2021)
[30]ProQuest (2021), *ProQuest Dissertations & Theses Global,* https://search.proquest.com/pqdtglobal (Accessed 11/5/2021)
[31]CORE (2021), *CORE API v2,* https://core.ac.uk/searchAssets/docs/#!/articles/searchArticlesBatch (Accessed 11/5/2021)
[32]Wikidata (2021), *Joining the ERM: Core executive decision-making in the UK, 1979-1990,* https://www.wikidata.org/wiki/Q106600328 (Accessed 11/5/2021)
[33]Poulter, Martin (2017), *A step forward in the sharing of open data about theses.* Bodleian Digital Library, 19 Jul 2017, http://blogs.bodleian.ox.ac.uk/digital/2017/07/19/a-step-forward-in-the-sharing-of-open-data-about-theses/ (Accessed 11/5/2021)
[34]Wikidata (2021), *LSE Theses in Wikidata,* https://w.wiki/vAK (Accessed 11/5/2021)
[35]Wikipedia (2021), *Wikipedia: list of infoboxes,* https://en.wikipedia.org/wiki/Wikipedia:List_of_infoboxes (Accessed 11/5/2021)

Images: CC0 (1.0), Wikidata, 3dman: Pixabay , OpenRefine,

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:

Karen Pierce: PierceKF@Cardiff.ac.uk

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

Book reviews: Please contact the editors.

ISSN 2399-9667

MDG website: cilip.org.uk/cig

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

MDG twitter: @CILIPCIG

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