

Catalogue and Index

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



Editorial

December 2018, Issue 193

Welcome to the December 2018 issue of Catalogue and Index, which brings you papers from the 2018 CIG conference. The issue contains 17 articles based on papers from this conference, followed by a conference report. This is truly a bumper issue of Catalogue and Index!

The papers from the conference are wide-ranging in their subjects. However, at their heart, they all share a common purpose: how to enhance metadata. Among other things, the papers demonstrate the resourcefulness of the current cataloguing world and the appetite (and need) for upskilling in order to carry out metadata work. Another common theme is the collaborative nature of the work, showing how the success of these cataloguer-led initiatives enhancement of the metadata often involves collaborations between cataloguing specialists at different institutions, cataloguing and IT specialists, cataloguing and academic staff, cataloguing and other metadata specialists from the broader information world, cataloguing staff and metadata providers, and many more types of collaborations.

The issue starts with an article by Jane Daniels which describes the collaborative LMS project in Wales (part of WHELP), and the associated developments and projects in collaborative cataloguing which have resulted from this work. Concetta La Spada describes her work with metadata at Cambridge University Press, and its involvement in the Metadata2020 project. Alan Danskin discusses the British Library's metadata strategy, and developments relating to upcoming systems changes, and their strategy relating major changes such as FAST and 3R. Andrea Del Cornò writes about using RDA to catalogue a collection of ephemera at The London Library, illustrating the challenges of using RDA for non-modern material. Sally Rimmer gives an account of a project at the University of Derby which sees repository metadata added to the library catalogue, and the consequences of this addition.



Contents

- 1-2 Editorial
- 3-8 Jane Daniels
- 9-17 Concetta La Spada
- 18-27 Alan Danskin
- 28-37 Andrea Del Corno
- 38-40 Sally Rimmer
- 41-44 Rachael Lamme
- 45-50 Penny Dougeris
- 51-55 Helen Williams
- 56-69 Will Peadar
- 70-81 Martin Kelleher
- 82-86 Joshua Barton
- 87-93 Louise Howlett
- 94-100 Anastasia Kerameos
- 101-106 Amy Staniforth
- 107-109 Jane Daniels
- 110-115 Casey Cheney
- 116-118 Frances Marsh (review)

Penny Doulgeris describes the move to a new LMS at the International Atomic Energy Agency, and the challenges that this presented especially in moving to MARC21 metadata. Helen Williams discusses a project at LSE to add records for blog posts to their institutional repository, including the complications of planning a project where many quantities are unknown and how the project had a positive impact on the relationship between the library and academic departments.

Will Peadon describes his various metadata initiatives at Aston University, including some detailed accounts of using MarcEdit for tasks such as authority control, reclassification and metadata enhancement. Martin Kelleher describes the process and technical details of using an EBSCO service to automate the provision of library metadata for their journal and other collections. Joshua Barton discusses some of the ethical issues of cataloguing zines, outlining how this brings new ethical responsibilities for cataloguers towards zine authors. Louise Howlett discusses developments in serial standards especially in recent years and the work of the ISSN community, including the changes wrought by LRM and the efforts to ensure serials are carefully considered in general bibliographic standards. Anastasia Kerameos describes the British Film Institute's work in standardising serials cataloguing. Amy Staniforth writes about a survey by the Jisc Data Community Group related to the National Bibliographic Knowledgebase (NBK), and shares their analysis and recommendations. Jane Daniels then outlines an idea which would see NBK also include catalogue records for resources which were lost and/or withdrawn, with a view to providing a "metadata memory" of the UK's collections. Casey Cheney describes the work by Backstage Library Works in converting AACR2 records to RDA, and to prepare catalogue records for the eventual move to Linked Open Data (LOD). Finally, Frances Marsh provides a report from attending the CIG 2018 conference.

We hope that the articles in this issue illuminate the wonderful work of the UK and international cataloguing communities.

Karen F. Pierce: PierceKF@Cardiff.ac.uk

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

What is the WHELP Shared LMS Consortium?

We are a subset of libraries that are members of the Welsh Higher Education Libraries Forum. In 2015 we went out to joint tender for a new library management system and discovery layer and by October 2016 we had all migrated to Ex Libris Alma & Primo. Our Consortium comprises all 9 HE institutions in Wales, the National Library of Wales and the Welsh National Health Service libraries. We also collaborate with the Open University in Wales because the OU also uses Alma & Primo. We number approximately 36.5 FTE (latest figures Jan 2017) dedicated cataloguers working with the assistance of Acquisitions, E-Resources & Systems colleagues. 22.5 of these FTES are employed at the National Library of Wales. What we lack in numbers we make up for in ambition and productivity especially now that we can collaborate with our colleagues across WHELP.

Previously at CIG2016

Sian Thomas² (National Library of Wales), Dr Amy Staniforth (Aberystwyth), and I presented a paper on our collective cataloguing experience using Alma at Conference in September 2016.

At that point, with the final group of WHELP libraries still to go live in October 2016, we had held our first Cataloguers' meeting including a briefing from Ex Libris on the creation of a shared catalogue and shared cataloguing workflows using a Network Zone; surveyed and recorded numbers of cataloguing staff and specific expertise in cataloguing subjects, formats or languages; identified the cataloguing standards in use and authority control policies; recorded the bibliographic record suppliers, catalogue enrichment services and cataloguing tools used across the institutions.

We had also established a Cataloguers Group on Yammer (a social network tool) to post documents for consultation and comment and pose questions & seek advice; and embarked upon the task of lobbying for more & better quality Marc records for the Alma knowledgebase or [Community Zone](#).³

2nd Cataloguers' Meeting Gregynog Hall March 2017

We held another meeting of Cataloguers and Systems colleagues at Gregynog Hall in March 2017. We met to compare and contrast our institutional records in preparation for creating a set of shared RDA cataloguing templates. We wanted to drive up the quality of our original cataloguing and inform legacy data enhancement. We invited Bernadette O'Reilly, Bibliographic Standards Librarian at the Bodleian Library, to talk to us about the shared cataloguing experience at Oxford. Prior to the meeting we all shared some record sets for various formats with each other.

1. Dumas, Alexandre (1998) *The Three Musketeers* (trans. by William Barrow) London: D. Campbell.

2. Thomas, S., Staniforth, A., Daniels, J. (2016) Alma: a cataloguer's experience, *Catalogue and Index*, 185, pp.12-16.

3. Alma Community Zone. Available at: <https://www.exlibrisgroup.com/products/alma-library-services-platform/community-zone/> [Accessed 24/11/2018]

We also shared them with Bernadette who used [MARC Report](#)⁴ to analyse them and give us some feedback. My colleague, Dr Amy Staniforth,⁵ presented in more detail on the templates at this years' CILIP Wales Conference and her presentation will be available shortly.

Outputs

We rigorously and systematically reviewed the sample records. Our differing cataloguing practices were discussed and recorded and by the end of the event we felt emboldened to sign up to create basic RDA templates for comment by our colleagues. Post-its were used to catch ideas for collaborative cataloguing and training & consortial purchase of cataloguing tools etc. We then themed the post-its and produced a report, with actions points. It's this document that has largely informed our progress to date.

Our subsequent activities & accomplishments can be summarised under 4 headings:

Shared Catalogue & Cataloguing

One of the major benefits that WHELFF have been looking to realise is a shared catalogue & shared cataloguing. This is a facility offered by Ex Libris in the form of a [Network Zone](#)⁶ so after our first Cataloguers meeting in 2015 we started investigating the creation of one. It became apparent that it was a large and potentially expensive process because, unlike the other consortia using a Network Zone, we did not have a pre-existing merged catalogue. We realised that we would need to secure the services of a third party to work with us to reduce the chances of inappropriate match and merge when trying to create a single database of all of the WHELFF records; and that there would be additional on-going costs to staff a central office to administer the NZ.

As we were still trying to learn how to use Alma and Primo consolidating our knowledge of the new LMS and discovery layer became the priority. Additionally, there was no appetite for further financial outlay. In 2018 there is renewed interest in a shared catalogue with the WHELFF Management Board asking for a review of the options and recommendations. The driver is the possibility of consortial management of eResources. We already have collaborative negotiated procurement deals for electronic resources managed via [WHEEL](#)⁷ (Wales Higher Education Electronic Library.) There is duplicated effort across the Consortium managing our respective eResources (often the same packages and deals) in Alma.

Could we unify to simplify? We are now carrying out a second round of consultations with consortia around the World who have Network Zones; ORBIS-CASCADE and Cal-Tech in the United States; LIBSYS in Belgium; BIBSYS in Norway. But a Network Zone is no longer the only option.

4. MARC Report Software. Available at: <http://www.marcofquality.com/soft/softindex.html> [Accessed 24/11/2018]

5. Staniforth, Amy. (2018) "A template for the future": Developing shared practice with colleagues across Wales. *CILIP Cymru Wales Conference, Aberystwyth 10th May*.

6. Introduction to the Network Zone. [https://knowledge.exlibrisgroup.com/Alma/Product_Documentation/010Alma_Online_Help_\(English\)/100Managing_Multiple_Institutions_Using_a_Network_Zone/01_Overview/01_Overview_of_Collaborative_Networks](https://knowledge.exlibrisgroup.com/Alma/Product_Documentation/010Alma_Online_Help_(English)/100Managing_Multiple_Institutions_Using_a_Network_Zone/01_Overview/01_Overview_of_Collaborative_Networks) [Accessed: 24/11/8/2018]

7. WHEEL (Wales Higher Education Electronic Library) Available at: <http://whelf.ac.uk/whelf-sub-groups/wheel/> [Accessed 24/11/2018]

The Jisc [National Bibliographic Knowledgebase](#)⁸ (NBK) has aroused considerable interest and support here in the UK. The NBK will succeed COPAC, a union catalogue of research and specialist libraries, and will include physical and eResources records from libraries to facilitate the exposure and management of collections across the UK. All of the WHELF institutions have contributed their records to Jisc for COPAC and the NBK and we are also active participants in the [Jisc NBK Community Data Groups](#).⁹ These groups have been established to look at a range of issues around library bibliographic and holdings data and to advise on possible interventions to drive up data quality and produce more efficient workflows. Myself and Dr Amy Staniforth are members of the Metadata Group; Sian Thomas contributes to the Tools group; and Gareth Owen to the Systems Group. Another possible NBK service could be the sharing of bibliographic records, initially original records created by libraries. There has to be further investigation of the licencing models for the re-use of vendor records. The library bibliographic data ecosystem is complex so Jisc have employed Ken Chad Consulting Ltd to review current practices and report back. WHELF awaits the findings with interest as we use a variety of methods to source records for copy cataloguing.

Purchasing Cataloguing Services/Tools

After the Gregynog event we successfully lobbied WHEEL for a consortial subscription to the RDA Toolkit, essential for the development of our RDA Templates. We have not pursued a consortial purchase of Marc Report although we definitely still need to analyse our records for errors and overall quality. But perhaps there are now other ways of doing this. We could use MarcEdit and Open Refine and Jisc are in development with elastic search functionality for the NBK. This would give NBK contributors a way to visualise their data and identify metadata quality issues.

Cataloguing Standards

The core RDA templates are already in use across the Consortium. Institutions then apply their own cataloguing policy to enhance the core records to suit their user group's requirements. We have signed off templates for monographs, DVDs, databases, theses, and junior fiction with templates for serials and junior non-fiction in review. Exhibition catalogues, audio recordings, maps, atlases, and artists books templates are in production.

We have also started mapping the use of local tags 9XX across the Consortium. Standardising the use of these would enable further configuration of a cross-WHELF search in Primo resulting in a superior search and discovery experience.

One of the reasons for selecting the Ex Libris Products was the promise of automated workflows for eResources. Ex Libris libraries can dispense with loading vendor records for eBooks and databases locally, opting instead to use the records in the CZ. Unfortunately the quality of the record is inconsistent and sometimes records are missing. Although the Company makes it easy for you to report errors or omissions there can be delays sourcing metadata. Setting aside those aggregators who for legal reasons cannot contribute their records to the CZ e.g. because of licence restrictions around re-use of records, there **are** publishers and content providers who **are** at liberty to share.

8. National Bibliographic Knowledgebase. Available at: <https://www.jisc.ac.uk/rd/projects/national-bibliographic-knowledgebase> [Accessed 24/11/2018]

9. Jisc National Bibliographic Knowledgebase Community Data Groups. Available at: <https://blog.ccm.copac.jisc.ac.uk/2018/04/27/national-bibliographic-knowledgebase-community-data-groups/> [Accessed 24/11/2018]

As we were having the same conversations about metadata issues, often with the same content vendors or publishers, we decided to work together to produce a generic template outlining the business case for good quality record contribution to the CZ. The template can be adapted for use with any Company. The Cataloguers illustrate the impact on search and discovery in Primo if records are poor or non-existent. A few screenshots from Primo is usually sufficient. The Business Case template was approved for use in November 2017 and there have definitely been improvements in the CZ as more companies have opted to contribute Marc records for their content but it is an ongoing project.

Training

We did consider consortial RDA Toolkit training but with the news of the redesign to accommodate the [IFLA Library Reference Model \(LRM\)](#)¹⁰ and the beta launch of the new Toolkit this June we decided to wait until the changes had been implemented and consolidated. In the meantime having to use the Toolkit to create the templates has been a great piece of CPD. The feedback and questions generated when templates are posted for comment are a fantastic way of sharing knowledge and skills. It also provides an opportunity to review institutional cataloguing policies.

Artists' Books

We have been lucky enough to get WHELF Staff Development funding for two events in 2018. In February this year Dr Karen Pierce, a cataloguer from Cardiff University, organised an [Artists' Books Cataloguing day](#)¹¹ at her institution. These resources present certain cataloguing challenges so having experts like Sarah Bodman and Maria White provide training for us on this genre and then practice cataloguing some titles was invaluable. We know that artists' books are present in collections across Wales; and an artists' books RDA cataloguing template has been promised off the back of this event.

MarcEdit and OpenRefine

Then in March we had a [MarcEdit & OpenRefine training day](#)¹² at the National Library of Wales in Aberystwyth. These open source packages are fantastic free tools for all kinds of metadata assessment and wrangling. The training was delivered by Owen Stephens and was pitched as an introductory level event. It attracted cataloguers and systems staff alike. Not everyone was a beginner but having colleagues in groups of mixed ability was very useful. We could share past experiences and ideas for future metadata projects in a friendly, supportive and informal atmosphere. We hope to re-convene in 2019 and have a "show and tell" session to share our practical experiences.

10. IFLA Library Reference Model (LRM) <https://www.ifla.org/publications/node/11412> [Accessed: 24/11/2018]

11. Pierce, Karen F. (2018) Artists' books cataloguing day. WHELF Wales Higher Education Libraries Forum. Available at: <http://whelf.ac.uk/page/2/> [Accessed: 24/11/2018]

12. MarcEdit & OpenRefine training day WHELF Wales Higher Education Libraries Forum. Available at: <http://whelf.ac.uk/marcedit-openrefine-training-day/> [Accessed: 19/8/2018]

Workforce Development

The issue of workforce development is definitely something that we need to address and some of our cataloguers already design and offer cataloguing training within the Consortium. Dr Karen Pierce¹³ trains staff working in the NHS libraries in Wales and presented on this at the CILIP Wales Conference this year in Aberystwyth.

WHELF Staff Development Fund 2018/19

We know that creating, sharing and maintaining metadata has never been more important at the local, regional, national or indeed, global level. For all of us the challenge is to ensure that we have a skilled cataloguing workforce now and in the future. We are very lucky that we have the opportunity to apply to the [WHELF Staff Development Fund](#)¹⁴ to run events for our Group. The 2018/19 budget is now available so we need to identify our training requirements and bid for funds to satisfy them.

Projects for 2019

Our WHELF Shared LMS Project Officer will present a Report on the options for a shared catalogue to Management Board in the next 3 months. The success and impact of the NBK, which potentially opens up all kinds of possibilities for sharing metadata not just in Wales but nationally in the UK, will have an effect on the decision-making. We hope to initiate and complete a feasibility study for the establishment of a NACO Funnel for Wales. This idea was discussed at Gregynog in relation to metadata standards but a UK Funnel has been on the radar since 2012 when Dr Deborah Lee¹⁵ presented on the topic at the CIG conference that year. Currently in Wales only our National Library can send Name authority records to the Library of Congress for addition to the Name Authority file. If more cataloguers in Wales could create and contribute records to the international NACO programme we would be able to establish access points for creators of works in our local and special collections and increase exposure of our collections via linked open data initiatives.

Jisc have now received all of the WHELF holdings to load into COPAC and the NBK. Once the loading is complete we would like to use CCM Tools to compare and contrast our collections. We need to know what resources we have in common; where do the rare and unique collections reside within Wales; what collaborative cataloguing opportunities are there for creating or improving records; can we work across institutional boundaries to improve metadata for the benefit of all?

We still need to review consortial copy cataloguing workflows. At the moment we have a mixture of institutions using a variety of sources and vendors and different acquisitions workflows ranging from the very basic to very advanced.

13. Pierce, Karen F. (2018) Cataloguing with the AWHILES libraries: training, sharing, supporting and dealing with the obscure. CILIP Cymru Wales Conference, Aberystwyth 9th May. https://www.cilip.org.uk/members/group_content_view.asp?group=200145&id=736742 [Accessed: 24/11/2018]

14. Help us create a new WHELF Staff Development Fund. Available at: <http://whelf.ac.uk/help-us-create-a-new-whelf-staff-development-programme/> [Accessed: 24/11/2018]

15. Lee, Deborah. (2012) Collaborative authorities: introducing the UK NACO funnel project. *Catalogue and Index*, 169, pp. 30-36.

Obviously with the national discussion started by Jisc¹⁶ around the licensing of vendor records for copy cataloguing we are keen to contribute our experience of the problems and inefficiencies that can ensue as well as our ideas for resolving them.

We will await Ken Chads' report with interest but are committed ultimately to using services assessed on the basis of record quality, cost, widest consortial subject coverage and interoperability with Alma. Many challenges and opportunities lie ahead but perhaps another quotation from *The Three Musketeers* will sum up the dynamic universe that WHELP cataloguers and anyone working with metadata will recognise. "*Never fear quarrels but seek hazardous adventures!*"

We will face these dynamic times together and ultimately what will make our dream work will be the team work.¹⁷

Biography

Jane Daniels is Bibliographical Librarian at Cardiff Metropolitan University. In addition to cataloguing in Higher Education Jane has been a Subject Librarian; part of a technical services team in a public library authority; and a local government information officer. She started her career working in medical libraries. She is a graduate of the College of Librarianship Wales in Aberystwyth.

This paper was written by Jane Daniels on behalf of the WHELP Shared LMS Consortium Cataloguers & Systems Staff. The WHELP Shared LMS Consortium Cataloguers are a group of cataloguers and systems staff living and working in Wales and using the Ex Libris Alma and Primo products.

jdaniels@cardiffmet.ac.uk @JaneDaniels1

16. Ruddock, Bethan. (2018) The library bibliographic data market: rights, responsibilities, and restrictions. Transforming our library support services. Available at: <https://libraryservices.jiscinvolve.org/wp/2018/08/the-library-bibliographic-data-market-rights-responsibilities-and-restrictions/> [Accessed: 24/11/2018]

17. Maxwell John C. (2002) Teamwork makes the Dream Work, Nashville, Tennessee, J. Countryman.

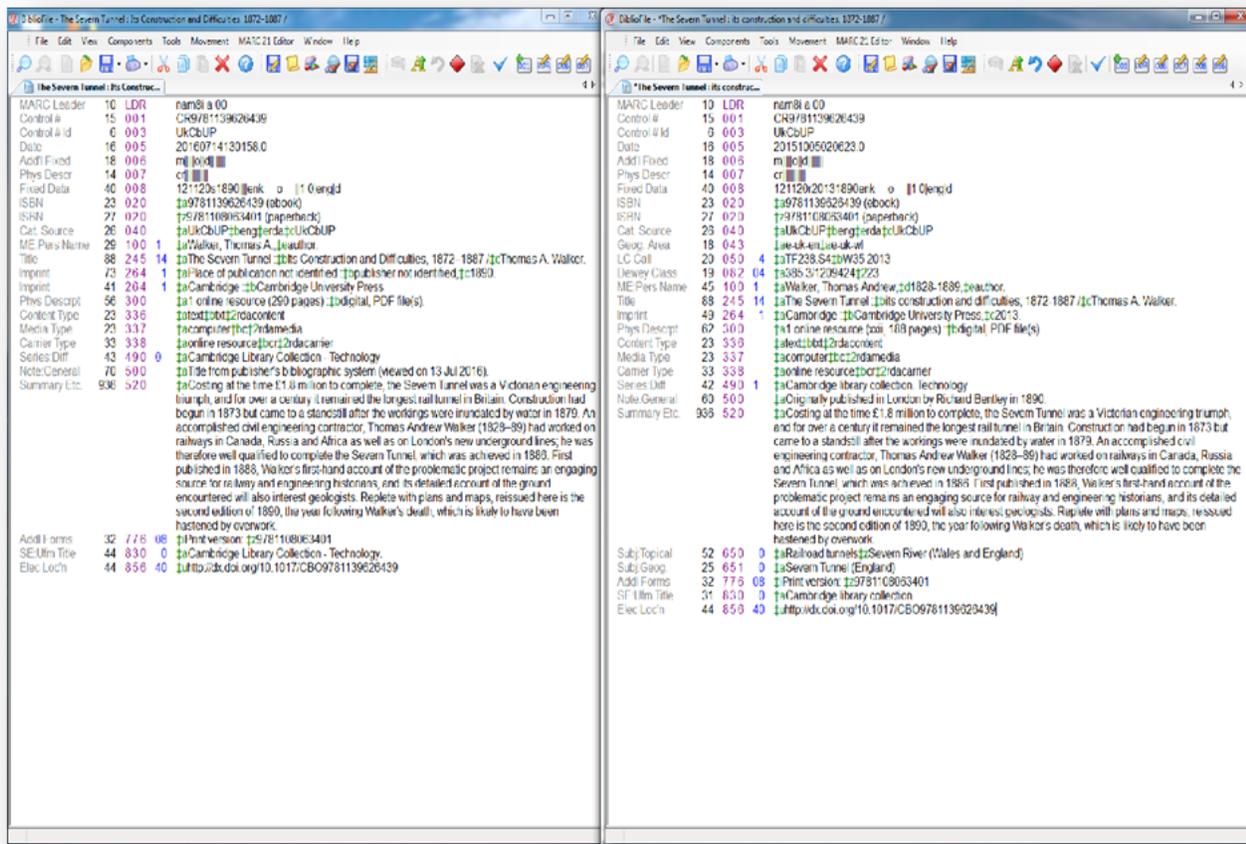
What do you think when you hear the word Revolution? You would probably think about one of the most famous in history, the French Revolution of 1789.



You could also think about the Industrial Revolution, which from England spread during the 19th century and changed the world.

Since we are cataloguers and librarians, I think that our minds go long, long ago, to a dark room where amanuensis friars copied books and preserved those for posterity. And then one thousand years later, in 1439 a new revolution changed books and reading forever, the invention of the Printing Press by Johannes Gutenberg.





On the left there is the automated record, with many good elements, on the right there is the full record, which contains many more elements, and in particular, subject headings. These are fundamental for the best discovery of resources, since subject searching remains one of the first approaches in the use of online catalogues among users. Using my knowledge of programs such as MarcEdit and Bibliofile, I re-catalogued the majority of our eBooks, about 28.000 titles.

Since the majority of our print titles are catalogued by the Library of Congress (LC), and these print records are then available online for free, I have downloaded all these records from the LC catalogue and other free sources, and then merged them with our automated records.

Obviously merging them was only the beginning, since there were many things to check before declaring those records ready to use. For example, many of these titles were originally catalogued according to the AACR and AACR2 so, since we wanted to supply RDA records, the data needed to be updated in order to conform to the new cataloguing rules. Also, over the years the Library of Congress had discontinued many of its classification numbers and also changed formats within its subject headings, so this data needed to be updated. And let's not forget human errors, like putting a print ISBN where it was not supposed to be, which got me records with two 245 fields for two completely different titles!

Yes, it has been quite an experience, but the results are worthwhile. Now these full records supplied freely by CUP (anyone can download them from our platform) contain:

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

Also, in these records that require them,

- Language and Geographical codes (041 and 043 fields)
- LC Authorized forms of Series' titles (490 and 830 fields)
- Corporate or Conference main entry (110 and 111 fields)
- Title variants (130, 240 and 246 fields)

The process of producing metadata has changed in such a way that nowadays when a new collection is being developed, its metadata is decided at the beginning of the process. At CUP, metadata is not something that is developed once the products are ready to be published, but is developed at the same time; this is to ensure that discoverability is immediate and not compromised by delays.

Every month, I enhance records for all CUP new titles and for our partner publisher titles.

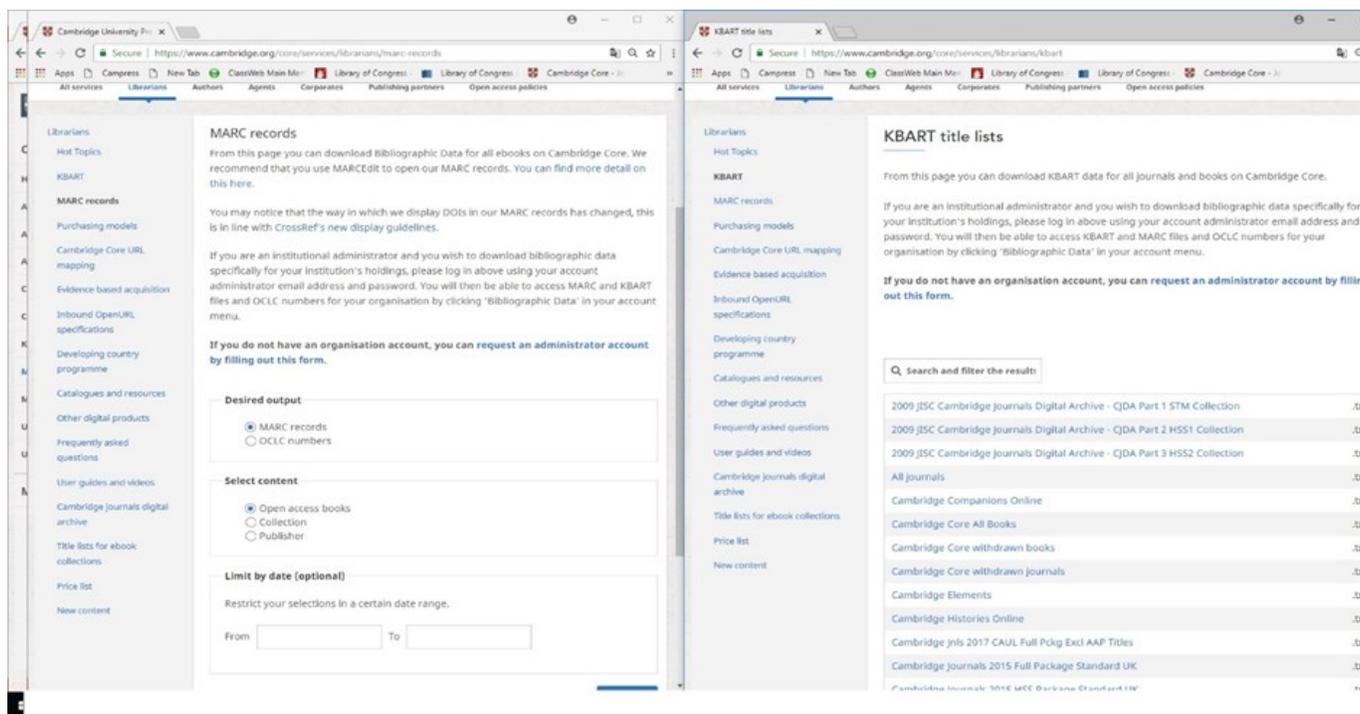


Once I complete enhancing the records, they are loaded onto Cambridge Core, our platform.

Cambridge Core: eBook MARC records and KBART lists

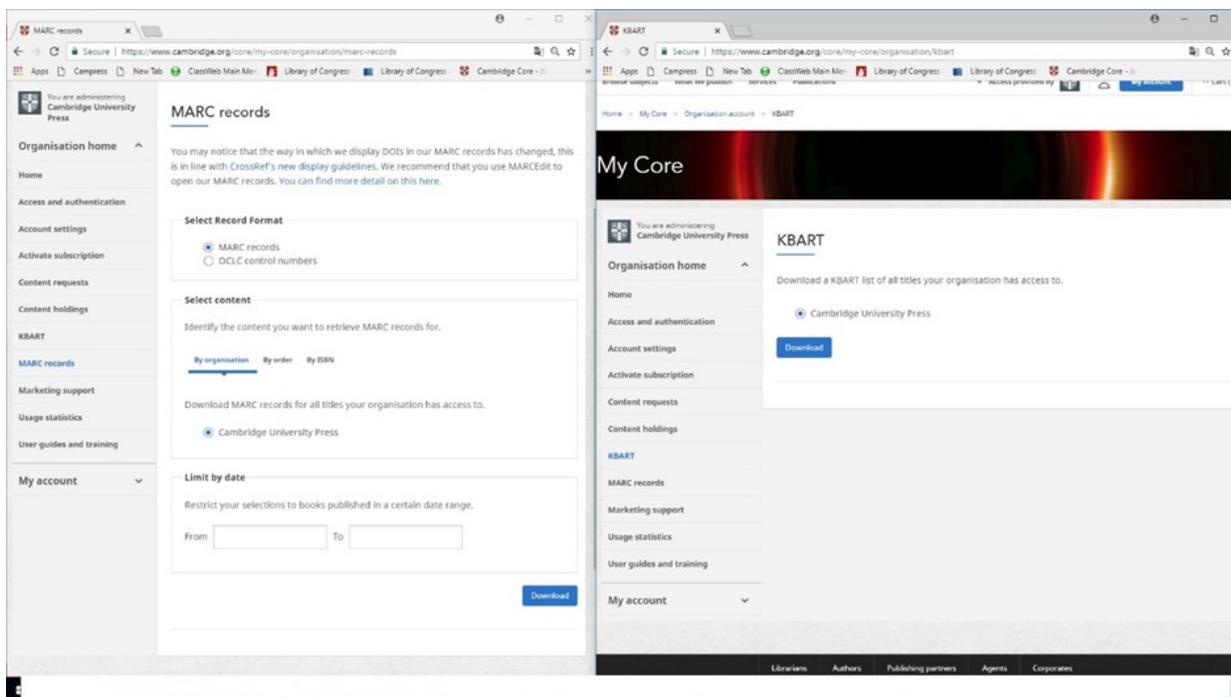
Librarians can download MARC records and KBART lists for free in two ways.

From the front end



On the page on the left, they can refine their search by content and by date. On the page on the right they can search within the page for the KBART list they are looking for. This page contain both eBooks and e-journals KBART lists.

Or, if they have an organization account on Core, they can log in and access to the MARC records and KBART lists areas.



On the MARC records area, they can download records by Organization, by Order and ISBN and refine their download by date of publication.

On the KBART lists area they can download a KBART list of their holdings.

A special project: MRS Online Proceedings Archive MARC records



Something that has always distinguished CUP has been is great care for customers' needs.

In the past, many customers had express interest in MARC records describing the MRS Online Proceedings Library Archive, which is a journal but in its print form is a collection of monographs.

In 2017 I had created full records describing this product but still data for about 400 titles was missing. So the only solution was to go and find the monographs themselves and that's what we did.

We contacted the British Library and asked to be allowed to have access to these titles and explained our purpose. We were the first publisher to formulate such a request and we definitely surprised them! We visited their office at Boston Spa and got the data we needed.

And the result, was full records for this collection.

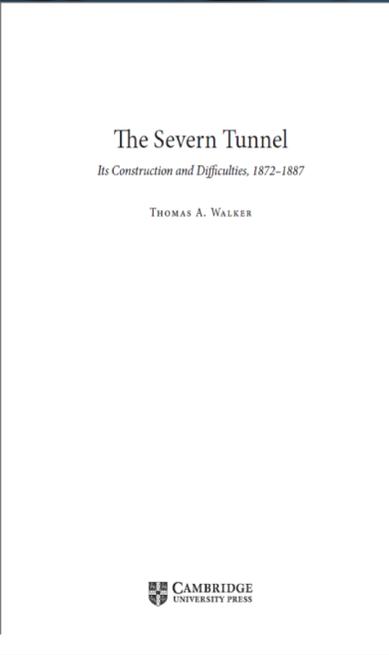
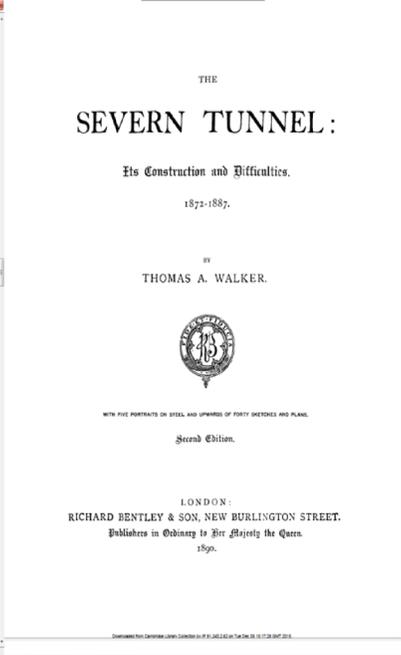
Special project 2: the Cambridge Library Collection MARC records

A second project on which I have been working is to create full records for the Cambridge Library Collection.

The Cambridge Library Collection (CLC) is a beautiful collection of books, with titles not only in English but also in French, German, Italian, Latin and other languages. These titles were all originally published between the 16th and the 20th century, so they are all out of copyright, and they have been brought back to life for the use of students and researchers.

The collection was born from a joint project between CUP and the University of Cambridge. It's a collection of significant historical interest with which we have achieved amazing results in terms of use and sales. But we know that we could achieve better results if the eBook MARC records supplied for this collection were improved.

CLC eBook Marc records

Title Page	Title Page verso	Original Title Page
		

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi, Mexico City

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9781108063401

© in this compilation Cambridge University Press 2013

This edition first published 1890
This digitally printed version 2013

ISBN 978-1-108-06340-1 Paperback

This book reproduces the text of the original edition. The content and language reflect the beliefs, practices and terminology of their time, and have not been updated.

Cambridge University Press wishes to make clear that the book, unless originally published by Cambridge, is not being republished by, in association or collaboration with, or with the endorsement or approval of, the original publisher or its successors in title.

The original edition of this book contains a number of colour plates, which have been reproduced in black and white. Colour versions of these images can be found online at www.cambridge.org/9781108063401

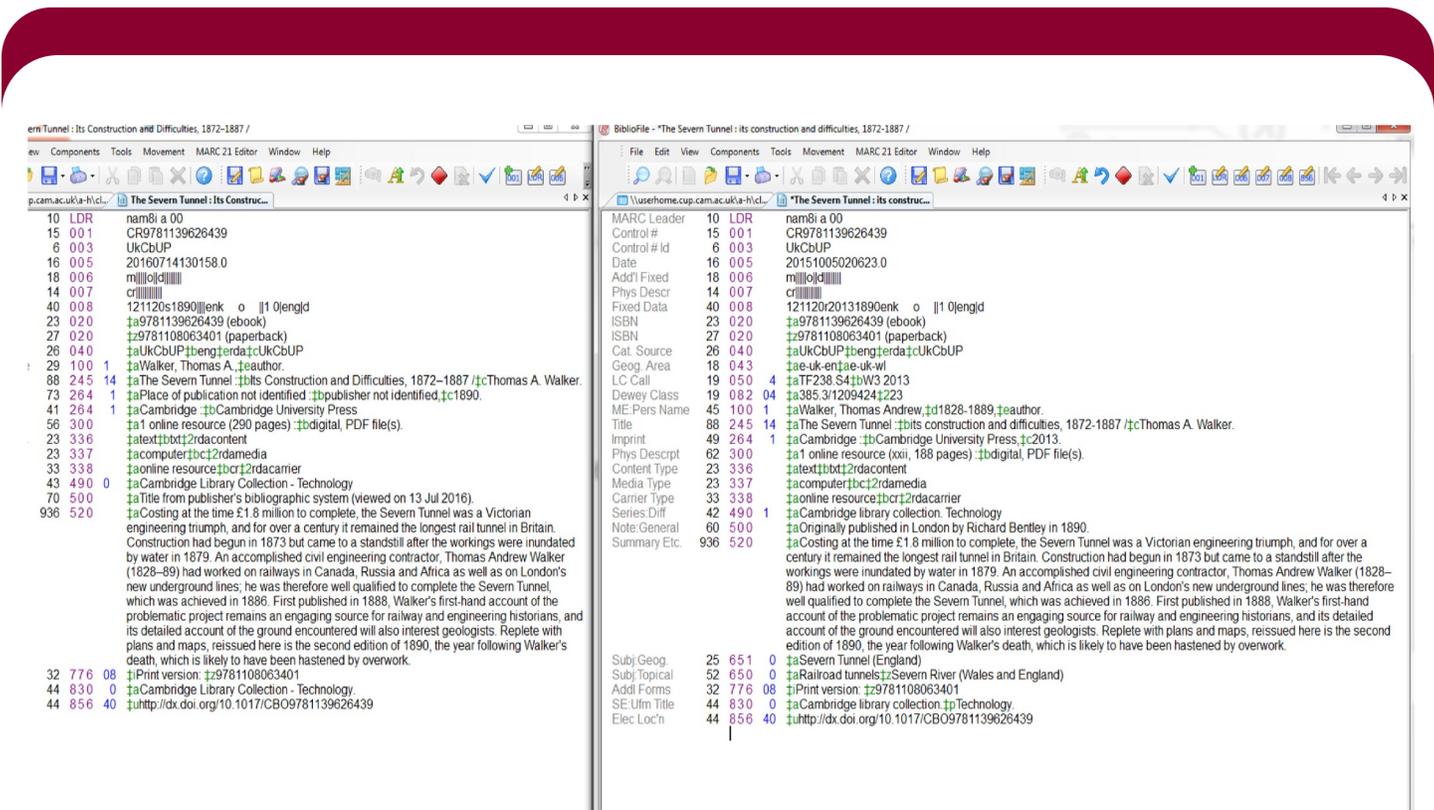
Original pub. date 1890
Print Publication date 2013

THE
SEVERN TUNNEL :
His Construction and Difficulties,
1872-1887.
BY
THOMAS A. WALKER.

WITH FIVE PORTRAYS ON STEEL AND UPWARDS OF FORTY ENGRAVINGS AND PLANS.
Second Edition.
LONDON :
RICHARD BENTLEY & SON, NEW BURLINGTON STREET.
Publishers in Ordinary to Her Majesty the Queen.
1890.

CAMBRIDGE UNIVERSITY PRESS

In this image you can see the structure of these eBooks. There is a new title page, a new title page verso, containing the new and the original print publication date, and a reprint of the original title page. Unfortunately, our system doesn't have the data about the original publisher and place of publication, so our records look like the example on the left (see below): with "Place of publication not identified" and "Publisher not identified" in the 264 field (elements that are of no help at all for users in their searches); also there are no subject headings; no classification numbers and no other useful access points.



On the right there is the same record but upgraded with all the correct access points and this is the kind of record we want to supply for this collection. The process of building these improved records has taken me a bit longer than it did for the CUP ebooks, because we don't have LC print records for this collection.

So I need to copy (after checking that everything is still correct) what I find in the LC records for the original publication. However, I will still need to check each eBook because many of these titles are two or more titles put together for the first time, so they have never been catalogued together before and it will be necessary to catalogue them from scratch.

As you can see this task is too onerous for one person so we have recently employed an external agency, Skope Knowledge Center, and we aim to have the completed records on our platform and delivered to Third Parties in early 2019.

CUP and the Metadata2020 Project

Metadata 2020 is a collaboration that advocates richer, connected, and reusable, open metadata for all research outputs, which will advance scholarly pursuits for the benefit of society.

In 2017, the Cambridge University Press became one of the participants in Metadata2020. It has given us the opportunity to make our contribution to what the Project is trying to achieve.

As a publisher, we understand that processes need to change. That ideas and solutions do not have to come up only when something is not working or it needs to be changed, for example, from librarians' requests, but directly from publishers and service providers. Only by working together can we supply better metadata. We want to advocate this to other university publishers and we hope that more of them will join us in this project.

CUP and Third Parties

Every month we supply metadata (MARC records, XML metadata, KBART lists), to Third Parties, like ProQuest, Exlibris, EBSCO and OCLC, to cite the major ones.

We want to make sure that the metadata that our customers will receive is correct and fit for purpose. We try to mediate between our customers and the Third Parties in case of any issues and we aim also to stipulate better agreements so that both sides are satisfied.

In 2017 we signed an agreement with Exlibris to have our enhanced MARC records added to the the Alma Community Zone. The Alma Community Zone is a great project, sharing cataloguing is a great way to promote collaboration among libraries; unfortunately a problem with this is that it is not always possible to guarantee the quality of the metadata within it. It is for this reason that we have signed the above mentioned agreement, so that we can guarantee that our metadata in Alma is of good quality.

With OCLC we have signed a similar agreement so that our enhanced MARC records and KBART lists are used to populate their Knowledgebase. Also, we have started to have monthly meetings with ProQuest/ Exlibris and OCLC where we discuss issues that customers have brought to our attention.

If there are any issues, we encourage customers to, raise tickets with the Third Parties and give us those tickets' numbers so that we can use them in our meetings and correspondence with the Third Parties.

Conclusion

I thank the CIG committee for giving me the opportunity to be one of the speakers at the CIG Conference 2018 and also for offering me a role within the committee. I am sure it will be a rewarding experience for me and I hope that what I contribute will be of help in developing current and future projects of the CIG Group.

Biography

A native of Sicily in Italy, Concetta completed her BA in Conservation of Cultural Heritage in 2007 at the University of Messina. During her university years she worked in various libraries and cultural institutions in Italy. She completed her MA in Archival and Library Science at the University of Rome 'La Sapienza' in 2009. In 2011 she moved to the UK where she worked as a Cataloguer for Blackwell, Baker & Taylor and YBP/ EBSCO, acquiring great experience in cataloguing books and ebooks. In September 2015 she moved to Cambridge University Press where, as Senior Library Data Analyst, she uses her knowledge of cataloguing and metadata to improve the quality of the metadata supplied to customers and third parties. From 2017 she is the Cambridge University Press representative on the Metadata2020 project.

Over the years the Library's collection processing workflows have gradually accrued standards, systems and policies adapted to processing large quantities of physical collection items. This has created a series of silos that are reasonably efficient in their own terms, but which are not directly interoperable; don't support discovery across the collection as a whole and are not optimised for digital resources. An increasing amount of time and effort is spent cutting against the grain to find "work-arounds" and fixes that will enable us to maintain services with systems, standards and processes that are rooted in the past. The world is changing faster than our infrastructure. What I am mainly going to talk about is how our metadata strategy is adapting to meet the challenge of realigning our tools to meet these challenges.

I've spoken about the Collection Metadata strategy in this forum before, but as 2018 marks the conclusion of the current phase, I want to start with an update of where we have got to. This is the vision we set out in 2015:

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

So, we still have a couple of years to go before attaining metadata Nirvana. The strategy set out three main objectives for the Library's collection metadata:

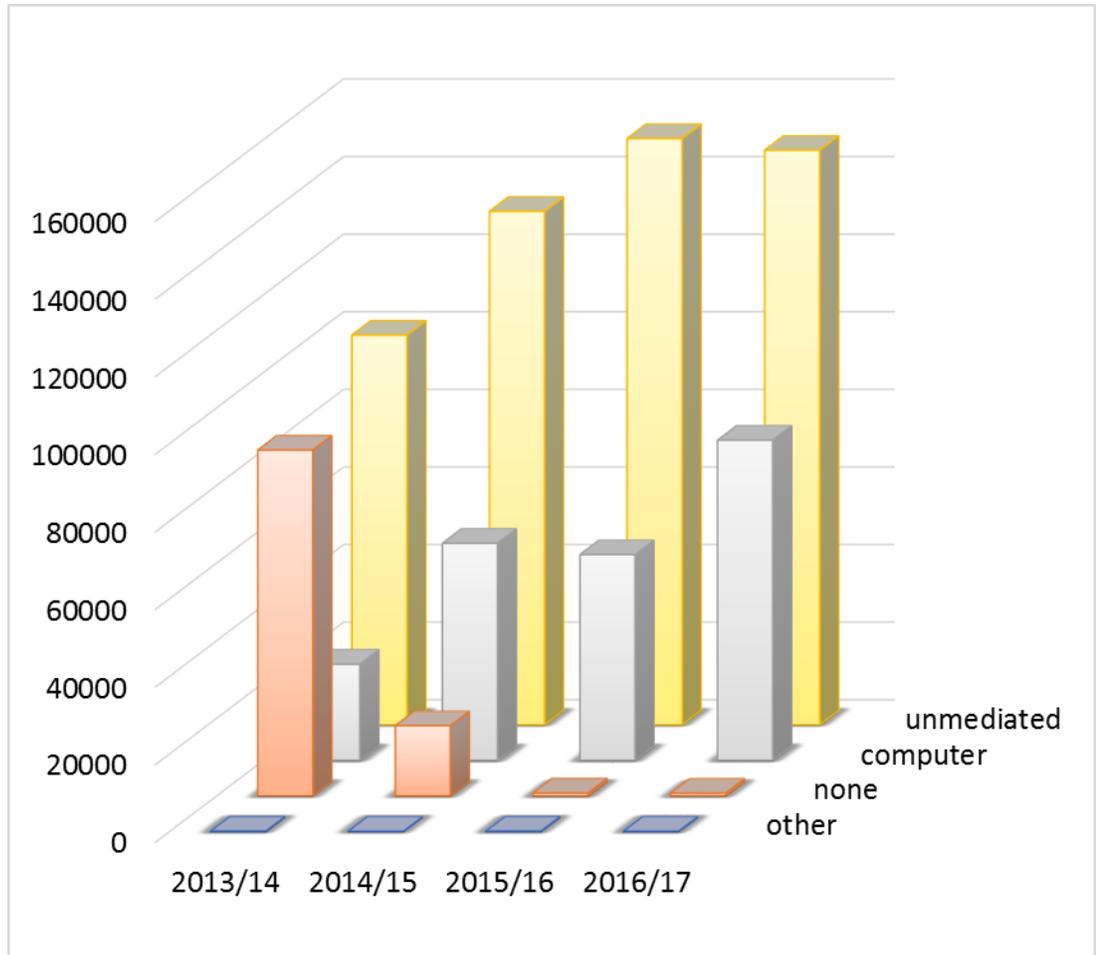
- 1. Drive efficiencies in the creation, management and exploitation of collection metadata to support delivery of the Library's strategic priorities and programmes**
- 2. Improve the Library's return on investment in its collection metadata assets by ensuring their long term value is maintained for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

In addition, we also have to support and enable the Library's strategic purposes which are articulated in *Living Knowledge*, and in several major portfolios which I'll come back to a bit later.

Driving efficiencies in the creation, management and exploitation of collection metadata

This histogram which follows is from the annual audit of data in our Aleph production database. It illustrates the values assigned to the media type element, for items processed since 2013. It shows the growth in the number of resources that require a computer to access the content. The number of items that are unmediated, i.e. can be used without any device or equipment, levels off.

These changes reflect the extension of legal deposit, in 2013, to include non-print publications. In 2013-2014 only 12% of the intake required a computer for use; 47% was unmediated, i.e. did not require any kind of device – except possibly glasses or contact lenses - for use. However almost as many did not contain the relevant metadata. As you may recall, media type was introduced to MARC to support RDA and RDA was implemented in 2013, so this tranche included a lot of AACR2 records.



We implemented some fixes to our imports that ensured content, media and carrier type would be assigned automatically to copy cataloguing. This process used algorithms that were tested and implemented on legacy data.

The figures from 2016/17 show that 36% of the intake processed now requires a computer for use. The intake increased from less than 25K in 2013/4, to over 82 thousand in 2016/17. Printed books and journals are holding steady, so that represents a very substantial increase in intake. There was no matching increase in cataloguers, so Collection Metadata was asked to develop an automated solution.

This process focuses on the auto matching and upgrading of the records accompanying Legal Deposit eBook. It is a development of a Batch upgrade process originally developed to enhance CIP records automatically.

(1) Publishers deposit content and metadata by FTP. The digital item is ingested to the Digital Library Store. The metadata comes to Collection Metadata usually in ONIX, sometimes in a spreadsheet.

(2) The first step is to transform the data from ONIX, etc. to MARC for load to our systems. There are two current versions of ONIX and different implementations. We write and maintain multiple transformations using XSLT. The overall objective is to merge the publisher data with richer library descriptions. This is done by successively matching the publisher records against selected data sets.

(3) Match with Cataloguing in Publication (CIP) data to enrich basic publisher records for dissemination to the other legal deposit libraries and our catalogue. In this instance we are retaining the CIP record and updating it with current publication details from the publisher metadata. CIP accounts for about 25% of the e-book intake. The remaining 75% of records require further matching

(4) Match with existing good quality OCLC eBook records. If there is a match, the OCLC record overwrites most of the publisher record. If there is no match on e-books:

(5) Match on records for print publications. If there is a match, the OCLC record overwrites the Work and Expression (content specific) attributes, but not the Manifestation (product specific) data.

(6) Unmatched records remain in the system until the point when they are statistically unlikely to find a match. These are reported for cataloguing to the Digital Processing team, who also resolve any exceptions identified earlier in the process.

Quality is a fundamental issue. Prior to the introduction of the upgrade process in late 2017, there was an 8% gap in quality between our un-upgraded (ONIX) e-book records and the overall quality of cataloguing. Batch upgrade has reduced this gap to 2% which is attributable in part to latency in the system.

The automated process currently handles 70% of the e-book intake without any cataloguer intervention. Only 6% of the intake is currently being processed manually; 13% is still in process. However 11% of intake has been through the batch upgrade process without finding a match. The relative immaturity of the process and the volatility of e-book publishing mean that it is not yet clear whether this represents a fundamental gap in capacity or a short term bulge. The failures are subject to a continuous improvement process to determine why they fail and what can be done to reduce the failure rate. Matching is heavily reliant on ISBN and one line of research has revealed that different ISBNs may be assigned to identical products for different markets, e.g. “public”; “discount stores” and “libraries” may all have copies of the same book, but it each will have a different ISBN. We are developing tools to cluster “sibling” ISBNs to overcome this problem.

On average the process generates 1,511 records per week or roughly equivalent to 8 extra cataloguers. Immediately following conversion to MARC and ingest of the content the publisher records are added to our production base and are available in the catalogue. These are interim records, which are adequate to find a known item, but whose discoverability is limited by the lack of controlled terms to provide context, including name authority control, DDC, LCSH.

The process was the outcome of a lot of hard work by expert staff and was dependent on the long term investment in tools to manipulate MARC data at scale and in procurement of datasets with permissive licences for re-use.

Improving the Library’s return on investment in its collection metadata assets

The relative quality of our metadata varies over time and poses many challenges, for example:

- attributes that are taken for granted in current cataloguing practice are not populated in many retrospectively converted records and their absence inhibits discovery and diminishes the reliability of result sets;
- the granularity of collective descriptions is inadequate when the “collections” are atomized for digitization.

The Library had catalogue records for the volumes or folders in which the art works were contained, but we would be receiving digital images of each work, so more granular metadata would be needed. We were given the pdfs for the published catalogues and asked if we could generate MARC records. Plausible looking MARC records were delivered within about three days and after further refinement and review a methodology was developed that can be applied to each of the 6 volumes, when they are received. These analytical records for the physical images can then be cloned and recoded for the digital images when we receive them. As the curator said, *“it really does look great and it will already have saved about a year of my time.”*

Opening up more of the Library’s collection metadata

Opening up the collection metadata means removing barriers to access where possible. Our free Z39.50 service enables selective download of our MARC records. Users are asked to register and accept our terms and conditions. The service is very widely used and popular, as can be seen from the map below.



Z39.50 2015	Z39.50 2018	Variance
1307 Users	1965 Users	+658
111 Countries	125 Countries	+14

Researchers who want to query bulk data have access to an increasing number of data sets in our Researcher Format. Researcher Format enables download of data sets in csv (comma separate values) a simple tabular form that can be used with Excel or OpenRefine for data analysis and cleaning.

Researcher Format 2015	Researcher Format 2018	Variance
6 Data Sets	18 Data Sets	+12
0 Updated	18 Updated	+18

These Researcher Datasets, related to specific exhibitions, themes or collections are regularly updated and freely available for download under CC0 licence. We also provide bespoke data sets for which we usually charge.

For example, the Graphic History Company asked for a list of the authors and titles for all the works ever published by Hachette or any of its imprints. We provided something like this:

93949	Thirion, Andr��	Revolutionaries without revolution	1976
93950	Thirkell, Angela	August Folly : A Virago Modern Classic	2014
93951	Thirkell, Angela	Wild strawberries	2012
93952	Thirlaway, Joanne	BTEC National travel & tourism	2007
93953	Thirlby, David	The Chain-Drive Fraser Nash. [With plates.]	1965
93954	Thirlby, G. F. (George Frederick)	L.S.E. essays on cost	1973
93955	Thirlby, G. F.	L.S.E. essays on cost. Edited by J. M. Buchanan and G. F. Thirlby	1973
93956	Thirlwell, John Bedford	Strength of Materials	1952

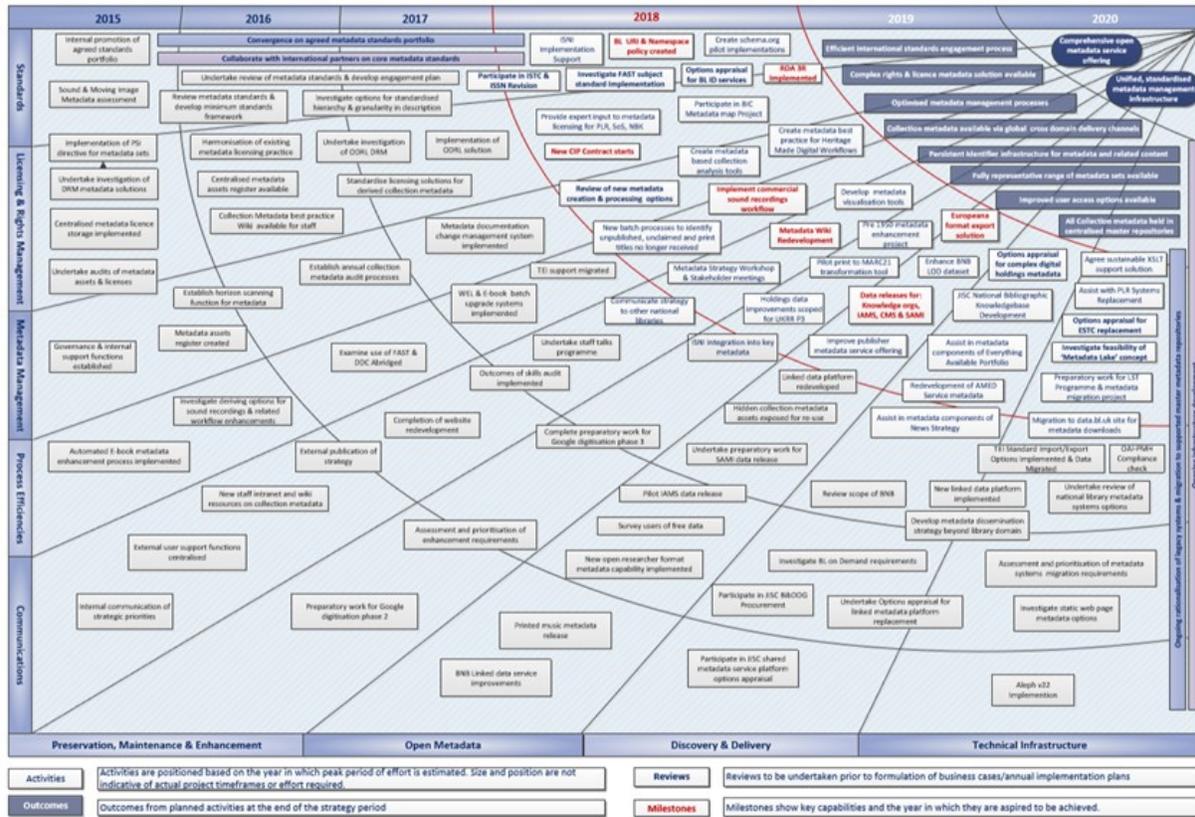
Which they used to create this...



River of Authors Image    Graphical History Company

The *River of Author's mural*, which is the centrepiece of Hachette's London HQ.

Evaluation



The Collection Metadata Road map 2015-18

A lot has been achieved...

- Significant efficiencies have been delivered, notably for e-books, but also by supporting “spreadsheet” input method to enable metadata creation by staff on projects and fixed term contracts for whom full Aleph training would not be economic
- Tools and other infrastructure developed to enrich and reuse metadata
- Our metadata is being more widely disseminated to more diverse audiences
- We have a much clearer understanding of our “hidden” and legacy metadata and remedial projects are underway to expose and enhance.
- Internal communication strategy has raised awareness and improved engagement

...but much more is being asked of us in 2018 than in 2015.

Looking ahead

Earlier this year we consulted on requirements for the next phase of the strategy. It came as no surprise that the need for convergence was a major theme. We still have too many workflows, too many standards and too many systems. This creates silos which create complexity resulting in the fragmentation of services, in particular discovery.

These silos are deeply rooted and can be traced back to the Library's creation in 1973. However, our three main cataloguing systems: for trade literature; for archives and manuscripts; for sound and moving images, are all nearing end of life. This provides a unique opportunity to re-model our system architecture, reduce the number of systems, converge standards, and change our business processes. These changes will also facilitate delivery by the Library's major strategic portfolios: Everything Available; Heritage Made Digital; and Everyone Engaged.

Library Systems Transformation Programme (LST)

The LST is a unique opportunity to refresh the systems architecture and simplify processes.

Major systems reaching end of life:

- Digital Library System (DLS)
- Aleph Library Management System
- Integrated Archives and MSS System
- Sound & Moving Image Library Management System

The LST programme manages a series of phased Projects to replace systems and migrate metadata and digital content. With the exception of DAMPS, the project to replace the in house Digital Library System, which is in the procurement phase, the projects are in the early stages of development. An RFI was issued last year to explore the library system market and we are currently carrying out further market analysis with suppliers and implementers. However the systems we procure are only tools and LST's scope encompasses business processes and convergence of metadata standards.

Standards

RDA: Resource Description and Access and FAST (Faceted Application of Subject Terminology) both have an important part in the Library's plans for the future.

RDA 3R

The RDA Toolkit Restructure and Redesign Project (3R Project) is a major update of RDA's content and the toolkit. The Toolkit is moving to a completely new platform with redesign of both the user interface and "back office" providing for a new look and feel and a much more efficient process for managing updates and translations. 3R is also implementing the IFLA Library Reference Model (successor to the FR "family" of models) as part of the project. LRM introduces new entities, including Place and Timespan. The RDA Aggregates Working Group has developed clear guidance for describing complex resources, including serials. The RDA Steering Committee (RSC) is also using the opportunity to implement some major changes to further internationalise RDA and make it more hospitable to metadata from other cultural heritage sectors and communities. In future RDA will impose fewer restrictions on how data is recorded by providing multiple options for recording each element. This will make RDA more flexible to accommodate and reuse "non-RDA" metadata. 3R also encourages the use of controlled vocabularies and data from outwith RDA, subject to recording their provenance.

From the British Library's perspective, 3R will help us to address some of the challenges we face in reusing metadata from book trade and music industry aggregators. We also expect that it will facilitate convergence of standards and workflows for our time-based media and possibly also manuscripts.

3R Implementation cannot take place earlier than Q4 2019 and is probably more likely to be during the first half of 2020. In the context of our current Aleph system and MARC 21 we do not anticipate a significant impact on cataloguers, but substantial work will be necessary to revise documentation, including policies, workflows and training courses.

FAST

Faceted Application of Subject Terminology (FAST) is an OCLC research project to derive faceted subject terms from Library of Congress Subject Headings. In essence FAST uses the same vocabulary and semantics as LCSH, but with a much simpler syntax.

In 2016 the Library consulted with the community on whether it would support a switch from LCSH to FAST in our subject indexing. The reason was principally the low entry cost of applying FAST compared to LCSH. FAST training takes less time and needs less supervision, and staff become productive more quickly with FAST than LCSH.

The community's response was mixed. There was interest in the potential for efficiency, but there was also caution about OCLC's commitment to FAST. There were also concerns about quality of indexing and loss of specificity. We therefore limited our application of FAST to special projects and we undertook to discuss a sustainable future for FAST with OCLC. As a result of the engagement by the Library and other institutions, including several Ivy League universities, OCLC has committed to developing FAST as a service, which we expect to be launched in March 2019. A FAST Policy and Outreach Committee (FPOC) has been set up with OCLC liaisons and representatives from the UK, USA and Australia. The Co-Chairs are Judy Jeng and Alan Danskin. FPOC met with Library of Congress in October to understand the process by which LCSH is managed and maintained, and discuss the relationship of the new FAST service to PCC's SACO Subject Access Cooperative. The committee subsequently met with the OCLC staff responsible for developing and maintaining FAST, and is now considering how FAST could develop to meet the needs of its users. The Library will continue to extend its use of FAST to increase the range of collection materials that we subject index, but FAST will not replace LCSH for the British National Bibliography without further consultation with stakeholders.

Conclusion

The articulation of a metadata strategy has raised awareness of the importance of metadata as an institutional asset. It has also given us a clearer understanding of the challenges and opportunities to improve the efficiency and utility of our services and processes. Critically it has also equipped us with tools to help shape a future in which we will be going with the grain.

Links and references

Unlocking the value: the British Library's Collection Metadata Strategy, 2015-2018. <http://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2015-2018.pdf>

Collection Metadata Strategy Roadmap, 2015-2018 <http://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-roadmap-2015-2018.pdf>

Reimer, T. (2018). The once and future library: the role of the (national) library in supporting research. *Insights*, 31, 19. DOI: <http://doi.org/10.1629/uksg.409>

River of Authors <http://theghc.co/project.php?project=hachette-uk-a-river-of-authors>

IFLA Library Reference Model <https://www.ifla.org/publications/node/11412>

FAST <https://www.oclc.org/research/themes/data-science/fast.html>

The Print Collection of Cassiano dal Pozzo: Ceremonies, Costumes, Portraits and Drama by Mark McDonald, Part C.I of *The Paper Museum of Cassiano dal Pozzo: A Catalogue Raisonné* Royal Collection Trust 2017

Biography

Alan Danskin is the Collection Metadata Standards Manager at the British Library with responsibility for the Library's engagement, deployment and application of bibliographic standards. Alan is Co-Chair of the FAST Policy and Outreach Committee and the British Library's representative to the European RDA Interest Group (EURIG). He is a former member of the Joint Steering Committee for Development for RDA (now RDA Steering Committee), the IFLA Bibliography Section Committee, MARC Advisory Committee, and Permanent UNIMARC Committee. He also served on the CIG committee for many years.

Alan.Danskin@bl.uk

Most recently, in May 2018, the Rare Books and Special Collections Section of IFLA (International Federation of Library Associations and Institutions) circulated amongst libraries and other institutions a survey on the implementation of RDA in rare materials cataloguing.¹ The scope of the survey was to build and share an accurate picture of how libraries worldwide have responded, or are planning to respond, to the adoption of RDA and to the specific challenges of cataloguing rare materials and special collections with RDA. The findings are yet to be made public. It is significant, however, that IFLA felt it helpful to initiate a debate on an important issue which to-date remains unresolved and surrounded by uncertainty.

At present a lack of consistency among different libraries, both at national and international level, appears to be the norm, whilst RDA developments for the cataloguing of rare materials remain in need of further attention. The want of consensus among the library community on how to move forward has become a hindrance and binding decisions have yet to be reached and agreed upon. The RBMS Policy Statements – a set of RDA-compliant guidelines for cataloguing rare materials – which should have been published in the current year, have now been delayed and expected to be finalise not earlier than 2019. This date too, however, remains doubtful. Eventually, RBMS Policy Statements should replace DCRM(B) – the Descriptive Cataloguing of Rare Materials rules. The RBMS Policy Statements should provide a set of cataloguing instructions for rare materials compatible and in alignment with RDA's guidelines for early printed resources. The RBMS Bibliographic Standards Committee, the driving force behind forthcoming developments, somewhat contradictorily, states that it remains “*neutral regarding RDA, neither encouraging nor discouraging agencies regarding implementation of RDA-acceptable DCRM records*”.²

Francis Lapka (Yale Center for British Art) and Audrey Pearson (Beinecke Library, Yale University) remind us of three key points of significance for users of special collections:

- Users must be able to distinguish clearly among different manifestations, and among variations within a manifestation;
- Users must be able to perform most identification and selection tasks without direct access to the materials;
- Users must be able to investigate the physical processes that produced the material, as well as post-production alterations.³

Echoing Sir Anthony Panizzi's statement: “[*a*] reader may know the work he requires; he cannot be expected to know all the peculiarities of different editions; and this information he has a right to expect from the catalogues.”⁴

*This article is based on my own paper presented at the 2018 Cataloguing & Indexing Group Conference (University of Edinburgh, 5-7 September). I am most grateful to Dunia Garcia-Ontiveros, CIG Chair, for her original suggestion to submit a paper proposal.

1. <https://www.ifla.org/node/40422> [accessed 1 August 2018]

2. <http://rbms.info/dcrm/rda/#RBMSPS> [accessed 1 August 2018]

3. Francis Lapka and Audrey Pearson, 'RDA & Descriptive Cataloging of Rare Materials: Developing Policy Statements for Special Collections Resources', *Catalogue and Index*, 183 (2016), 5.

4. Anthony Panizzi, Response to question 9814, 'Mr. Panizzi to the Right Hon. Ellesmere – British Museum, January 29, 1848'. Commissioners Appointed to Inquire into the Constitution and Government of the British Museum. Appendix, London: Her Majesty's Stationery Office, 1850, Q. 9814 (May 17, 1849).

These principles, put forward in 1848 by the future Principal Librarian of the British Museum Library, retain their validity and remain unchanged.

The RDA definition of an early printed resource reads: “[a] manifestation manufactured using a hand-press method, before the advent of machine printing in the nineteenth century.”⁵ Chapter 1.12 in the RDA Toolkit (added April 2017) – in an abridged form – states that:

*[t]he instructions for the provision of more detailed descriptions for early printed resources are also applicable, selectively or in full, for any resource determined by an agency to require a more detailed description. This determination may be made for any reason including, but not limited to, age, rarity, fragility, or value as part of a collection.*⁶

This statement is applicable to all of the London Library’s special collections. The London Library – an independent institution dating from 1841 founded by Scottish historian and biographer Thomas Carlyle (1795-1881) – successfully transitioned to RDA from AACR2 in January 2015. The transition was across its holdings in their entirety, including special collections and rare materials. Library management and Bibliographic Services staff were directly involved with the various stages of adoption and implementation of RDA. This allowed a more seamless and effective transition process and a more enthusiastic training or re-training of staff.

A set of guidelines for cataloguing rare materials and antiquarian books with RDA has been put together for internal use and to facilitate cataloguers. These guidelines are based on the policy created, and kindly shared, by the University of Kent.⁷ All libraries ought to put together RDA cataloguing policies for special collections and rare materials to avoid variations and conflicting cataloguing practices amongst staff and to facilitate both access and use. At the London Library, a tailored and detailed policy has yet to be completed however. The cataloguing of the *Per Nozze* special collection, for which the guidelines were adapted further, represents a pilot project which could be rolled out for all of the Library’s special collections. The *Per Nozze* or epithalamia collection – in excess of 2,500 bibliographical units – dating from the sixteenth- to the twentieth-century, represents a discrete collection amongst the London Library’s rich historical holdings.⁸ *Per Nozze* publications, are literary compositions, in verse or prose, printed to celebrate weddings – a custom and a publishing *genre* almost entirely confined to Italy. The origins of *Nozze* or *Per Nozze* publications (literally ‘for a wedding’) might be traced back to classical times, reaching a height of popularity through the nineteenth-century, but almost disappearing during the course of the twentieth.⁹ Originally, these compositions or collections of writings, often by different authors, were printed to celebrate royal or noble weddings. The first printed such example, identified by Italian scholar Olga Pinto, is an incunable dating from 1484, which includes nine Latin *carmina* composed by the Venetian humanist Francesco Negri (1452-1523?) and printed at Padua, northern Italy.¹⁰

5. RDA Toolkit, Glossary <<http://access.rdatoolkit.org/>> [accessed 05.06.2018]

6. RDA Toolkit, 1.12 Early Printed Resources <<http://access.rdatoolkit.org/>> [accessed 05.06.2018]

7. On the implementation of RDA at the University of Kent, see Josie Caplehorne and Rachel Dickinson, ‘Something Old, Something New: Rare Books and RDA’ *Catalogue and Index*, 183 (2016), 15-20.

8. On the London Library’s *Per Nozze* Collection, see Andrea Del Cornò, ‘Una collezione di *Per Nozze*: prolegomeni ad uno studio’, *La Bibliofilia*, 115 (2013), no. 1, 53-62.

9. The tradition continued throughout the early decades of the twenty-first century. On the *Per Nozze* as a literary *genre*, see Antonino Musumeci, *La musa e mammona. L’uso borghese della parola nell’Ottocento italiano* (Ravenna: Longo, 1992).

10. Olga Pinto, *Nuptialia: saggio di bibliografia di scritti italiani pubblicati per nozze dal 1484 al 1799* (Florence: L. S. Olschki, 1971), p. 1.

As early as the end of the sixteenth-century, however, writings dedicated to wealthy, non-noble families began to appear and the Italian vernacular became the preferred language replacing Latin. Outside Italy, the London Library's collection of nuptial poems is a unique resource whose importance is comparable only to those *Per Nozze* collections held at the National Library of Florence and at the Berliner Stadtbibliothek.¹¹

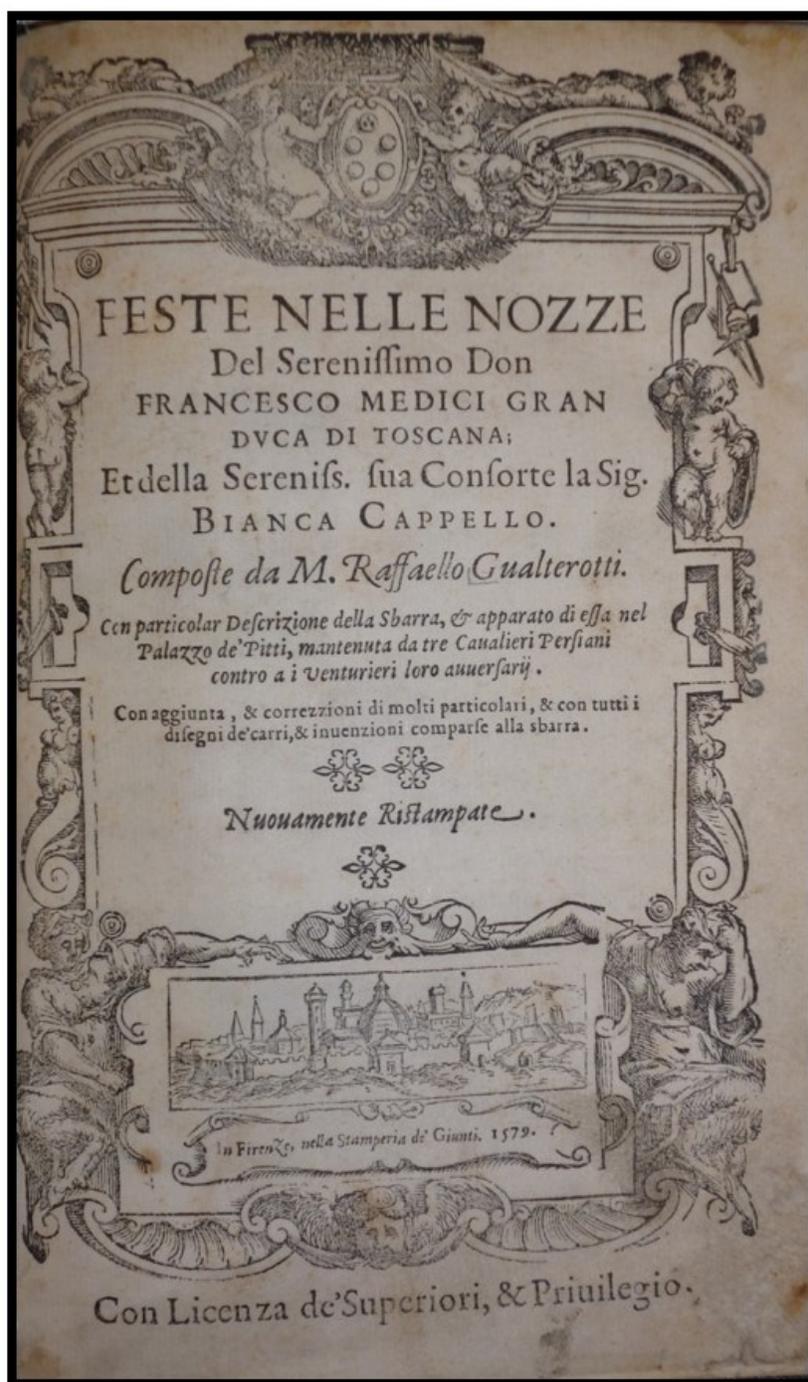
Per Nozze are by their very nature ephemera publications. Printed for specific occasions, often lacking the standard elements of conventional publications, *Per Nozze* appeared in limited editions, intended for private circulation rather than commercially sold. Whilst some have been preserved in libraries on account of their textual importance or fine bindings, a sizeable number were printed on poor quality paper, even on loose sheets, and by non-professional printers. These are now very rare or no longer extant. A great deal of planning was often required in the book-design and imposition. Title pages and frontispieces often became eye-catching *tableau-vivants* finely executed. The format varies through the ages, from *quarto*, or even *folio* and *grand folio* of early periods to more restrained *octavo*, *duodecimo* or even *sextodecimo* of the nineteenth- and twentieth-century.



©London Library, Nozze collection, no. 60.

11. These two collections are often referred to as the 'Miscellanea Capretta' and the 'Casella Collection' respectively.

The earliest *Per Nozze* held at the London Library is an elegant booklet in *quarto*, printed in Florence by the renowned printers Filippo and Jacopo Giunta, on the occasion of the wedding of the Grand Duke of Tuscany and lady Bianca Cappello in 1579.¹² The title page features elaborate decorative borders and, in a *cartouche* at the bottom of the page, a woodcut representing a view of the city of Florence. The corresponding London Library's record, created in-house, is reproduced below:



©London Library, Nozze collection, unnumbered.

12. London Library, Nozze collection, unnumbered.

Name Tag	Data
FMT	BK
LDR	00000nam 22002177i 450
001	000530302
003	OCoLC
005	20180820121429.0
008	960229s1579 it 000 0 ita d
040	a UkLoLIB b eng c UkLoLIB e rda e dcrmb
049	a L7LA
099	a Nozze
1001	a Gualterotti, Raffaello, d 1543-1638, e author.
24510	a Feste nelle nozze del serenissimo Don Francesco Medici Gran Duca di Toscana et della sereniss. sua consorte la Sig. Bianca Cappello : b con particolare descrizione della sbarra, & apparato di essa nel palazzo de' Pitti, mantenuta da tre cavalieri persiani contro a i venturieri loro avversarij / c composte da m. Raffaello Gualterotti.
250	a Nuovamento ristampate, con aggiunta, & correzioni di molti particolari, & con tutti i disegni de' carri, & invenzioni comparse alla sbarra.
264 1	a In Firenze : b Nella stamperia de' Giunti, c 1579.
300	a 58 pages, 2 unnumbered pages, 24 pages, 1 folded leaf : b illustrations (black and white) ; c 22 cm
336	a text 2 rdacontent
336	a still image 2 rdacontent
337	a unmediated 2 rdamedia
338	a volume 2 rdacarrier
500	a This volume is part of the London Library Nozze Collection.
500	a Gualterotti's poems "Vaghezze sopra Pratolino" and "Epitalamio nelle nozze della illustrissima signora Perregrina Cappello, et dello illustrissimo sig. conte Ulisse Bentivogli", pages 1-24 at end.
500	a Copy imperfect: plates lacking.
500	a The sbarra was performed in the courtyard of the Pitti Palace on October 14, 1579.

500	a The title page statement that the work is newly reprinted may refer to an earlier program for the spectacle without illustrations; see Mortimer, and also A.M. Nagler, Theatre festivals of the Medici, page 49.
500	a Dedication to the Grand Duchess by Filippo and Jacopo Giunti, A2.
500	a Woodcut title page border incorporating Medici arms at head and view of Florence at foot.
500	a Historiated and decorative initials.
500	a "Le invenzioni e disegni ... [delle] stampe sono del sig. Raffaello Gualterotti, intagliate da Accursio Baldi e Bastiano Marsili".
500	a "Tavola delle invenzioni, disegni, & mostre, di quelli che comparsero alla sbarra. Et prima" H2r.
500	a Signatures: A-G4 H2.
60000	a Francesco b I, c Grand-Duke of Tuscany, d 1541-1587 x Marriage.
60000	a Bianca, c Grand-Duchess, consort of Francesco I, Grand-Duke of Tuscany, d 1548-1587 x Marriage.
650 0	a Marriage z Italy v Gift books.
650 0	a Royal weddings z Italy z Florence v Early works to 1800.
650 0	a Tournaments, Medieval z Italy z Florence v Early works to 1800.
650 0	a Florence (Italy) x History y 1421-1737.
655 0	a Epithalamia 2 lcsh
655 7	a Nozze 2 London Library
7000	a Francesco b I, c Grand-Duke of Tuscany, d 1541-1587, e honouree.
7000	a Bianca, c Grand-Duchess, consort of Francesco I, Grand-Duke of Tuscany, d 1548-1587, e honouree.
70012	a Gualterotti, Raffaello, d 1543-1638. t Vaghezza sopra pratolino.
70012	a Gualterotti, Raffaello, d 1543-1638. t Epitalamio.
946	a rcp3167
979	a adc/ro

The MARC record leader is coded - as following ISBD – ‘i’ in LDR/18. The 040 MARC field (Cataloguing source) includes the subfield ‘e’ (Description convention), repeated as ‘erda’ and ‘edcumb’.

The title proper in the 245 MARC field is transcribed according to RDA 1.7 General Guidelines on Transcription. RDA permits a great deal of flexibility and it allows the transcription of information originating from elements elsewhere in the resource other than the title page, without the need to add an explanatory note. Often *Per Nozze* publications have ‘titles’ such as: *Nozze*, *Per Nozze* and other variants or combinations of the word *Nozze*. The title proper and statement of responsibility might not appear on the main title page, and the page normally at the beginning of the publication might not present the fullest information about the publication and the work or works it contains. These facts, often combined with lack of further standard bibliographical details, might cause some difficulties in identifying a specific manifestation of a work/expression.

With regard to inaccuracies, typographical errors and variant spellings in the title, which may occur with a certain frequency in early printed materials, RDA states that: “[w]hen instructed to transcribe an element as it appears on the source, transcribe an inaccuracy or a misspelled word unless the instructions for a specific element indicate otherwise”.¹³ Additionally, RDA only suggests to: “[m]ake a note correcting the inaccuracy if considered important for identification or access”,¹⁴ and: “[i]f the inaccuracy appears in a title and a corrected form of the title is considered important for identification or access, record a corrected form of the title as a variant title.”¹⁵ In these cases, the prescribed DCRM(B) instructions remain more satisfactory. The internationally recognised and accepted way of treating interpolations – enclosed within square brackets – although not ideal, immediately ‘flag up’ any discrepancies and alert the end user.

Interpolations in the transcription, such as *i.e.*, *sic*, enclosed in square brackets, and equally, interpolations for conjectural and unreadable text and for lacunae in imperfect copies might better serve the needs of the end users. It should be added that the use of these conventions remains widespread and accepted among cataloguers, curators, bibliographers, book dealers and antiquarian book collectors. Equally, the same can be said for the use of abbreviations, acronyms and Latinisms such as *s.n.* for ‘sine nomine’, *s.l.* for ‘sine loco’, and so on and so forth. Additionally, for rare books cataloguing, brevity presents the advantage of conveying information in a more direct and comprehensible way. Finally, with regard to the 245 MARC field, RDA is silent or indifferent to transposition and RDA guidelines give prominence to transcription over interpretation.

On the plus side, in RDA the 264 MARC field – which replaces the 260 one – is a repeatable field and allows a more precise labelling of the book production process, making a clear distinction between publisher, producer, manufacturer, etcetera. The 264 MARC fields used in conjunction with Catalyst – the London Library’s discovery system (Ex Libris Primo) – should allow Library members and other users to identify the information they need and the items they require much better. This in turn should limit unnecessary handling of special collection items and place less demand on the Library’s book-fetching services.

RDA states that: “[f]or early printed resources, distribution and manufacture statements relating to booksellers and printers may be treated as publication statements.”¹⁶ and data provided by the 264 MARC fields can be used effectively in combination with the information included in the 700, 710 and 752 MARC fields. The use of relator terms also provides helpful information. Appendix I of RDA includes an extensive, albeit not comprehensive, list of relationship designators for agents associated with a manifestation such as: ‘printer’, ‘engraver’, ‘previous owner’, etcetera.

In the case of *Per Nozze* publications, the ‘Added Entry-Hierarchical Place Name’, that is the 752 MARC field, may be of particular significance to scholars and researchers.

13. RDA Toolkit 1.7.9 Inaccuracies <<http://access.rdatoolkit.org/>> [accessed 05.06.2018]

14. Ibidem.

15. Ibidem.

16. RDA Toolkit 2.8.1.1 Publication Statement <<http://access.rdatoolkit.org/>> [accessed 05.06.2018]

Writings *Per Nozze* were usually privately printed in limited editions, often by local or obscure printers. Added entries for place of printing and the 710 MARC field for printers should therefore be added as these fields can provide valuable information and might be crucial in identifying a particular item. For this typology of publications, however, establishing the place of printing – and often equally a date – might entail a tiresome and time enveloping search.

Returning to the matter of the 264 MARC field, according to RDA, if more than one place of publication is named in the source of information, the record should indicate the place names in the order given by the sequence, layout, or typography of the names on the source of information. As exemplified in the following record, however, this does not take into account the fact that, for early printed material, especially for books printed in Italy, often the last place indicated on the resource is the proper place of publication:

```
245 10|aNuouo modo per imparar a leggere, e scriuere in due mesi. Et un breue Regola  
d'ortografia. Con un giuoco vertuoso. Di Bastiano Zanetti, il quale s'obliga con detta Regola  
insegnar a vecchi, & giouani  
264 |aIn Roma, Venetia, Bressa, Firenze, Verona, & di nuouo ristampata in Ferrara, per Vittorio  
Baldini  
300 |a4 unnumbered leaves;|c8o.  
500 |aSignature: [A]8  
710 2 |aBaldini, Vittorio,|eprinter  
752 |Italy|dFerrara
```

In the above example, therefore, the place of printing is Ferrara and not Rome, as correctly indicated in the 752 MARC field.

To record extent, in the 300 MARC field, RDA has introduced the use of a clumsy way of recording numbered and unnumbered pages, leaves and columns. The new form of syntax introduced by RDA to record pagination and foliation in resources issued as volumes remains unsatisfactory and places an unnecessary onus on cataloguers, often under pressure and operating within strict time constraints.

The RDA 'take what you see on the resource' approach and the elimination of abbreviations, should possibly be open to revision. Furthermore, in the case of collational formulae, according to RDA, these should be provided when considered important for the identification or selection of a manifestation. It is worth pointing out that a collational formula – a key element in rare book cataloguing and description – by definition is a symbolic alphanumeric (often including non-alphabetic characters) abbreviated representation of a physical object, a book, or, in RDA parlance, of a (printed) manifestation.¹⁷

RDA emphasises the importance and value of relationship designators. With regard to this specific *genre* of publication the relevance of relationship designators is self-explanatory. Publication of *Per Nozze* were often compiled or had their printing commissioned by a dedicator, not necessarily by the agent responsible for creating a work or, more prosaically, the author. At the London Library designators such as: 'honouree' (for the spouses), 'dedicator' and 'dedicatee', appended to subfield 'e' in the 700 MARC field, are applied. All 700 MARC fields are checked against the Library of Congress Authorities and authorised entries are created when necessary.

17. Similar concerns are equally raised in Adelaida Caro Martín and Roberto Gómez Prada, 'RDA and Rare Materials at the National Library of Spain', *Cataloging & Classification Quarterly*, 54:5-6 (2016), 305-316, and in Alberto Petrucciani, 'RDA: un'analisi critica alla luce della teoria e della pratica della catalogazione', *JLIS.it*, 7:2 (2016), 109-162 <<https://www.jlis.it/article/view/11784/10915>> [accessed 05.06.2018]

Often a detailed and careful reading of the preliminary pages is required to establish the exact nature of a relationship between a given name and a manifestation. In turn, this clearly necessitates specific linguistic skills and the ability to read and understand the text. During a pilot project, carried out over a period of two weeks, Bibliographic Services staff at the London Library, lacking a knowledge of Italian, had some perfectly understandable difficulties in establishing the exact relationship between a manifestation and the various personal names appearing in a given item. Around five-hundred and fifty records have been completed so far – the great proportion of which entailed original cataloguing – and they include the 655 *genre/form* MARC field. The 655 MARC field is a repeatable field and in addition to the Library of Congress prescribed terminology and controlled vocabulary – the heading ‘Epithalamia’ indicated by the value ‘0’ in the second indicator – all London Library’s *Per Nozze* records feature an additional 655 MARC field marked ‘Nozze’, with a second indicator ‘7’, source specified in subfield ‘2’, that is to say the London Library.¹⁸ On completion of the cataloguing project, all items within the *Per Nozze* collection will therefore be identifiable and retrievable using the 655 MARC field. To better serve the Library’s members and the wider research community, however, a ‘general note’ 500 MARC field with the text: “This volume is part of the London Library Nozze Collection” is added to all records and displays prominently.

Progress to-date in the cataloguing of the *Nozze* collection is a fine example of how specific challenges need addressing and how consistency in cataloguing practices for all special collections materials should be promoted and encouraged. RDA compliant records will enhance the discoverability and access to printed ephemera, remaining as part of valuable, but too often, hidden collections.¹⁹ Additionally, this will allow co-operation and resource sharing amongst libraries and similar institutions. Cataloguing printed ephemera, by their very nature, continues to be a challenge and RDA sometimes places additional onus on cataloguers.²⁰ Creating descriptive metadata and new RDA compliant records – almost invariably from scratch – is a labour-intensive and time-consuming activity.²¹

Overall, the experience at the London Library has been a positive one. Despite several difficulties and pitfalls – some foreseeable and others totally beyond planning - it is clear that at the London Library the correct decision was taken in fully embracing RDA. In 1841, Thomas Carlyle, the founding father of the London Library, convincingly stated that: “a *Library without a catalogue is Polyphemus without an eye on his head.*” It is undeniable that this statement retains its force of truth even more today. Libraries will need to allocate appropriate resources and skills to make their record future-proof and to maximize and fully exploit the benefits of RDA compliant records for end users, to further enhance global discoverability and promote best practice.

18. As an example, see the London Library’s record reproduced on pp. 32-33.

19. See Gordon Dunsire, *RDA (Resource Description and Access) and Its Application to Rare Books, Manuscripts, and Their Digital Surrogates* <<http://www.gordondunsire.com/pubs/docs/RDARareBooks.pdf>> [accessed 05.06.2018]

20. The definition of ‘ephemera’ remains a controversial issue, see Stephen J. Zietz, ‘Ephemera: MARC-Formatted Cataloging Records’, *Rare Books & Manuscript Librarianship*, vol. 7, no. 2 (1992), 152-159.

21. On the challenges and advisability of implementing RDA, see Michael Gorman ‘RDA: the Emperor’s New Code. A Brief Essay’, *JLIS.it*, 7:1 (2016), 99-107 <<https://www.jlis.it/article/view/11565/10748>> [accessed 05.06.2018]. See also Margaret Beecher Maurer and Roman S. Panchyshyn, ‘Understanding the Why: A Case Study in Managing the RDA Implementation’, *Cataloging & Classification Quarterly*, 52:3 (2014), 259-284.

Biography

Italian-educated, with a Laurea in Political Science from the University of Rome “La Sapienza”, completed his studies in London, specialising in rare books and early printed materials at the Institute of English Studies, University of London [MA in Librarianship, awarded 2004; MA in History of the Book, awarded 2012].

His published articles range over librarianship, history of printing and history of the book. An essay on the London Library’s collection of epithalamia, or publications *per nozze*, has featured in *La Bibliofilia* (no. 1; 2013). More recently, an essay on the *editio princeps* of Dante’s *Divine Comedy*, printed at Foligno in 1472, has been published by the Società Bibliografica Toscana (2014). A study of the English translation of Francesco Biondi’s *Eromena, or, Love and Revenge* printed in 1631 by prominent London bookseller Robert Allott awaits publication.

Since 2007 Andrea has been editor of the Italian Studies Library Group *Bulletin*, annual publication of the Italian Studies Library Group. He is presently Italian specialist at The London Library, where he has been in post since 2002.

Having your institutional repository records integrated into your library catalogue is acknowledged to be a good thing. Who wouldn't want all their institution's research outputs made visible and searchable alongside its other collections? However, costs to outsource this work can be prohibitive and it is often put aside as a "nice to have" facility.

So it was at the University of Derby until circumstances changed. The appointment of a Repository Librarian, a newly created post, and the impact of REF 2014 led to a greater awareness of our repository UDORA (University of Derby Online Research Archive). There was a subsequent increase in submissions further enhanced through promotional events showcasing the service but the impetus to really consider integrating UDORA within our library catalogue (Capita PRISM) and hence our discovery service (EBSCO EDS known at Derby as Library Plus) finally came after an enquiry from an academic who was nonplussed as to why the UDORA records weren't in the catalogue. The time was right to start a project to see if it could be done in-house.

From the start we wanted it to be feasible and sustainable for one part time cataloguer to pull into their existing workload and knew that, for maximum effect, UDORA records would preferably be searchable as a discrete collection within the library catalogue.

Preliminary work started by examining the metadata produced by our repository. Our repository is from Atmire¹ and the metadata scheme used is Dublin Core so the primary task was how to convert the Dublin Core to MARC data. There is a published crosswalk² and the first attempt consisted of downloading one record from UDORA with the aim to transpose the elements to MARC fields with a view to somehow automating this later. This was simple to do but of course with over 2,000 records in UDORA, and the ideas for somehow automating it later in short supply, manual retrospective transposition was not an option plus we had new submissions coming in all the time.

Further research led to OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) which is when the possible use of MarcEdit³ came to mind. MarcEdit, freely available, allows you to manipulate your metadata easily. Previously MarcEdit had only been used at Derby for editing e-book records prior to upload into the catalogue but it also has an OAI-PMH harvesting facility in its toolkit. Serious upskilling in MarcEdit was done by means of Youtube videos⁴ and a very comprehensive set of online guides produced by the University of Illinois⁵ which provided the knowledge to set about harvesting, converting and editing our repository data into usable Marc.

1. <http://www.openrepository.com/>

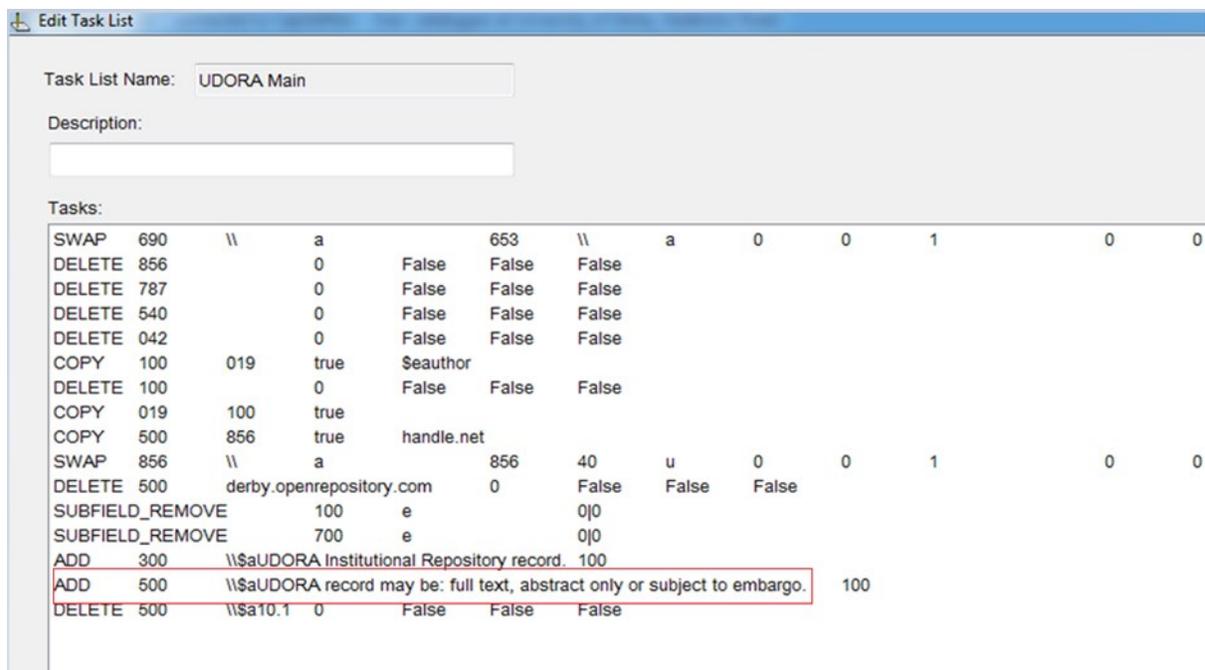
2. <https://www.loc.gov/marc/dccross.html>

3. <https://marcedit.reset.net/>

4. <https://www.youtube.com/user/tpreese/videos>

5. <http://guides.library.illinois.edu/MarcEdit>

We were then on to the concept proving stage. Marcedit allows you to download the whole repository, collections or a particular subset using a date range. The latter was chosen to produce a small test set. Marc came out but in a raw state which needed quite a bit of editing but this is quite simple as MarcEdit allows you to easily manipulate data by performing bulk editing using sequenced specific tasks. You can also save these sequenced tasks as task lists for repeated use. After trial and error our eventual task list was set at:



Some of the functions seen used here are swap, delete, copy, subfield remove, add. For instance we added a note to say that the record may be full text, abstract only or subject to embargo to clarify what the record is pointing to. We also deleted the relator subfields purely because in the first instance, for speed, we made them AACR2 records. Going forward we will use RDA and the RDA helper within MarcEdit to do so.

Our Library Management System (Capita Alto transitioning to Soprano) allows you to bulk import records to a particular profile which we used to import the test set. We set the profile up to enable us to run a management query which would provide the SQL needed to create a discrete UDORA collection in PRISM. We also needed a separate profile for our theses as we also wanted them to be picked up by the existing Theses and Dissertation collection in PRISM.

The concept was proved but how long would it take to do all the retrospective work and would it be feasible to fit new submissions into a regular upload schedule bearing in mind the time constraints of a part time cataloguer? Using MarcEdit to download the entire repository the sequenced tasks were applied with a few fields (001, 003, 008) added separately in bulk at the end. Using the MarcEdit split function we divided the approximately 2,300 records into batches of 100 to make it easier to manage. Each of these files were quickly checked and any errors amended. Each file took approximately 45 minutes to fully prepare and load including the manual tidying (which included coding subfields correctly and some series entry editing). The theses records were hived off to work on separately later as slightly different tasks needed to be applied. Over seven weeks the retrospective editing and importing of the 2,300 records took place as it fortunately coincided with a quiet period in the cataloguing year.

Work also began on formulating a process for monthly updates. MarcEdit does allow you to specify a date range to harvest from the repository but unfortunately in our instance it returned anything new or edited during the specified period.

However, it is relatively easy to tell by their repository handle number as to which are the most recent and you can sort records by field within MarcEdit.

Once we had finished loading all our existing repository records into the catalogue, and had a method for updating the new submissions, issues surrounding this upload became apparent. Specifically, the UDORA records were inconsistently skewing results in our discovery service and library catalogue. In the discovery service UDORA records would sometimes appear at the top of the results list ahead of full text subscriptions. This was problematic as the UDORA records may or may not lead to full text (depending on embargoes and other access restrictions). If the first link doesn't resolve to full text then there is always the risk of a student thinking that we don't have access when we might. Also we couldn't easily identify and only load repository records where we didn't already have a full text subscription and even if we could as subscriptions change quite frequently we couldn't keep amending the catalogue records to keep pace with them.

In our library catalogue (PRISM) a similar problem occurred. We have a results clustering facility enabled which groups together all formats and editions of the same title under the umbrella record of the latest version. Where the UDORA record was the latest version it could look like it was the only version available and if a student found that it didn't resolve to full text they might conclude that we didn't have full access to the work. Finally our attempts to create a discrete repository collection in our library catalogue were also thwarted by technical issues.

To solve the book clustering issue it was decided that we would purchase copies of books where our research community had had input and suppress the relevant UDORA records. This was a good news story to tell as well as solving this problem. Work on creating a discrete collection for UDORA items is still ongoing in collaboration with our IT services and library management system suppliers.

However, the discovery issue, once it became apparent that we couldn't fix it in the parameters of the discovery service, was more serious. Requests to see if the UDORA records could always be the last in a series of results was not possible as a subset of the catalogue could not be separated out from the rest of the catalogue records. So we took the decision to suppress all the repository records that were journal articles which left only the grey literature to be added to the catalogue and hence into our discovery service. We haven't given up on the idea of integrating the journal articles into our catalogue and will be working with our discovery service and repository providers as technology advances to progress this.

Positive outcomes for us include adding grey literature to the catalogue on a monthly basis and the future project of adding our e-theses. We are also investigating using MarcEdit to convert repository metadata into a format acceptable for use in an RSS feed on our university website.

The project raised some interesting, wider, discussion points in our library community. Library catalogues don't exist in isolation anymore. They are often connected to reading list services and discovery services which can be impacted, positively or adversely, by changes you make to your library catalogue. Then there is the nature of the catalogue itself, it traditionally points to full text and our repository records did not always resolve to full text. Whatever, the rich data we produce needs to work hard for us and tools such as MarcEdit can help us to this.

Biography

Sally Rimmer's role includes a remit to advise on the metadata aspects of the University of Derby's institutional repository (UDORA) and because of this she has become interested in converting and using data in multiple ways.

s.e.rimmer@derby.ac.uk

Rachael Lamme, Head of Community Outreach, Crossref

Metadata 2020 is a collaboration of publishers, librarians, service providers and providers of platforms and tools, data publishers and repositories, researchers and funders. They have the shared mission of improving the richness, connectedness, reusability and openness of metadata for all research outputs.

Richer, accurate, interoperable metadata is certainly a worthwhile goal, and one that would benefit many communities within the research space. There are many separate stakeholders who are working towards improving their metadata, however many of these efforts remained unconnected. Metadata2020 was launched to try to bring the work that these groups were doing together, and to make metadata improvements a top priority in areas where it may have languished as a 'nice to have' in the past.

The initiative was launched in 2017. It was conceived by Ginny Hendricks from Crossref. However, Metadata2020 needed to pull from the expertise of lots of different groups, so work started by bringing on board interested parties from the different communities. These parties formed interest groups in September 2017, and met to define the core problems for each of their communities. Today, around 120 people are currently involved between the different groups, which have now formed cross-community collaborative projects to address some of the core issues needed for metadata improvements.

These projects include:

- metadata mapping
- making metadata vocabulary more consistent
- forming best practice statements
- creating metadata evaluation tools
- forming and communicating business cases to enrich metadata
- communicating the importance of metadata to researchers

These projects are being undertaken by individuals with metadata experience from across scholarly communications, by individuals acting collectively to make improvements. The key to the group work is that they have been formed for a limited period of time to complete the work, and will then be reassessed to move the outputs of the groups forward, be that either via publicising the work that the group has done to the wider community, driving adoption or putting certain communities together via workshops, webinars or other means yet to be decided!

To come at the project from a Crossref perspective, we know that there are issues with publisher metadata management. It is costly for them, a huge challenge to implement for previously published material, they work with vendor systems whose requirements they have limited control over, and the community lacks an effective metadata distribution model. Publishers end up sending different subsets of metadata to lots of different end-points rather than being able to use one standard output. They often find it difficult to make the case to authors as to why good metadata (and supplying comprehensive information up front) is valuable to them.

Similar issues occur for service providers, platforms and tools. Metadata creators make assumptions that don't travel with the data, there is a lack of consistency in and between metadata schema, records can disappear without transparency as to where or why, there are no community standards about metadata vocabulary, and finally, metadata can easily get out of date if updates aren't provided by the metadata creators.



But there are opportunities too – the possibility of collaborating to find a consistent metadata vocabulary, working together to improve system workflows, developing business cases for improved metadata, mapping metadata to improve interoperability between services, service providers and publishers. Crossref has also launched [Participation Reports](#) for publishers so that they can easily see what metadata they're providing to Crossref, as sometimes this can be obscured by relationships with intermediaries or sit within one specific department. For example, publishers may be collecting references and ORCID iDs, but ensuring that information is in the metadata they distribute can be key to downstream discoverability and efficiency gains for researchers.

Those are just two communities struggling with common issues that they need to work together to solve, with metadata issues ranging far beyond their immediate constituencies. Hence the idea of forming groups and tasking them with finding solutions based on their shared knowledge.

At the CILIP Cataloguing and Indexing Group (CIG) conference in September 2018, the focus was on surfacing some of the key developments made in the metadata mapping and metadata vocabulary projects.

The role of the Metadata Recommendations and Element Mappings group is to converge communities and publishers towards a shared set of recommended metadata concepts with related mappings between those recommended concepts and elements in important dialects. It is a challenge for the community that there are many different ways that metadata is created, vetted, used and distributed; and the complexity of this makes finding new efficiencies and systems implementation difficult. Most groups face interoperability challenges with systems and processes, and there are silos within organizations themselves, making communications challenging.

The group was tasked with exploring solutions such as; identifying concepts included in relevant community and publisher recommendations, identifying concepts shared across recommendations and mapping across schemas to identify inefficiencies, breakages, etc. which may be leading to interoperability problems.

The group's work thus far has involved compiling an index of metadata schemas that different parts of the industry use. This list stretches to 36 different schemas from Dublin Core to BIBFRAME through PubMed and KBART. It's useful to have such a list to look at the scale of the task, but the decision was made not to try to map elements across all 36 schema, but just across the most popular/frequently used ones. Then, within these schema, the group listed out 34 elements and looked at how these are represented across each of the schemas.

Things that this exercise highlighted are that there are of course differences in how elements are named e.g. authors/contributors/creators, license_ref/rights/permissions/legal constraints/copyright permissions. The term 'abstract' was pretty consistent across most schemas, but absent in some, and there are lots of gaps! It's useful for this information to be exposed so that schemas and schema creators can identify these pitfalls, see where information might be going missing and adopt existing terms if adding to their own schemas rather than creating new ones.

Another project has been tasked with Defining the Terms We Use About Metadata. In order to communicate effectively about anything, a common language must be acknowledged, tacitly or purposefully. In the metadata space, there is not agreement on what words like property, term, concept, schema, title refer to, and different groups, librarians, publishers, researchers all use different vocabulary. This project specifically aims to create a glossary of words associated with metadata, both for core concepts and disciplinary areas, which should help accuracy when it comes to passing metadata through different workflows.

The first thing the group worked on was to list current definitions to find the most common uses of those definitions across the board. They are also working on a core metadata glossary of terms e.g. "concept", "schema", "title".

The end goal is to evolve a consistent core metadata glossary to speak to different research fields. It's important to acknowledge that there are other activities that are similar to this, for example CASRAI, the W3C group dealing with linked data and DCMI, all of whom the project is engaging or hoping to engage with.

The final project highlighted in the meeting is around creating shared best practice and principles around using metadata across the scholarly communication cycle, in order to facilitate interoperability and easier exchange of information and data across the stakeholders in the process. At the moment, there is a lack of central core principles, best practice and guidance, so lots of different stakeholders are creating their own (which don't necessarily match up).

The project group working on this have been working to define core principles for metadata around scholarly communications, created and disseminated in easily digestible ways for different groups. This also includes looking at issues around metadata ownership and governance.

The list is not meant to be comprehensive. If you think a resource is missing from this list, please let us know! Email us at info@metadata2020.org

Name (BPs = Best Practices)	Year (if known)	Primary Audience	Source type	General or type-specific?	Note
CHORUS Publisher Implementation Guide	2016	Publishers	Community Group	Journals/Conference Proceedings; Detailed	Provides recommendations for the data publishers submit about funding information for publications supported by publicly funded research. 2018 revisions expected
Cornell University Library Repository Principles and Strategies Handbook	2018	Libraries/Repositories	Library	Repositories	An organized and accessible description of different types of metadata created and used in library and archival settings
Crossref Books Advisory Group BPs	2016	Publishers	Community Group	Books; Broad	Provides descriptions of what the Crossref metadata repository requires to provide access and linking between books and journals. Some 2018 revisions expected
Crossref: Best Practices for Depositing Funding Data	2015	Publishers	Community Group	General; Detailed	General information for providing data for FundRef, the part of CrossRef that is a repository of funding sources for published research
Crossref: Depositing Reuse License Information		Publishers	Community Group	Journals/Conference Proceedings; Detailed	Describes CrossRef's policy for creating URIs with license information about access and use of a journal article or book; provides text and data mining users with a clear way of determining what they are permitted to do with content identified by a CrossRef DOI (digital object identifier)
D-Lib article: OpenDOAR Repositories and Metadata Practices	2015	Libraries/Repositories	Article	General; Broad	A study of how academic institutions create metadata for institutional repositories and archival systems; the scope of this study is restricted to DOAR (Directory of Open Access Repositories)
DateONE BPs		Environmental Sciences Researchers	Community Group	Data; Detailed; Environmental sciences	DateONE Best Practices database provides individuals with recommendations on how to effectively work with their data through all stages of the data lifecycle.
Emory University's Core Metadata Guidelines	2014	Libraries/Repositories	Library	General	Emory University Library's Core metadata items that represent the 18 minimum items are required to support search and discovery.
Journal Article Versions (JAV): Recommendations of the NISO/ALPSP JAV Technical Working Group	2008	Publishers/Libraries	Standards Body	Journals/Conference Proceedings; Detailed	These NISO/ALPSP Journal Article Versions (JAV) Technical Working Group recommendations provide a simple, practical way of describing the versions of scholarly journal articles that typically appear online before, during, and after formal journal publication.

Existing best practice documentation, available at:

<http://www.metadata2020.org/resources/metadata-best-practices/>

In late 2018, Metadata2020 ran two workshops, one in New York and one in London with participants from both the interest groups and the project groups. These workshops focused on working towards building a metadata flow diagram, to look at where metadata 'falls out' of the system and all of the different places it travels to. Work was also done to sketch out the 'big benefits' of metadata so that this can be used to communicate to businesses, researchers and other parties why it's worth investing time and effort in improving how we work with this valuable information.

Finally, the groups worked on best practice, breaking down the parties who work with metadata into a number of roles, depending on what they do (creator, curator, custodian and consumer) and look at what good practice would be at each point e.g. creators should provide the best possible metadata for the use case, curators should have quality control, custodians should make the metadata openly available - where possible - and metadata consumers should show the provenance of where they got their metadata from, especially if they are integrating it into their own tools and services. Note, an organisation can play more than one of these roles. This output is being shared with the wider group for discussion, and will then be released to the community more widely for feedback.

Now that the workshops are complete, the outputs will be taken back to the project groups to see what progress has been made and what the next steps should be. This might take the form of outreach around the group outputs, refocusing the group tasks or adding groups or individuals in whose voices are missing. It's important to note that the groups all consist of people volunteering their time to the project - it won't run forever, and the aim of the groups is to work for a short period of time on set projects to produce outputs - but sometimes if people are short on time, things will take a bit longer to come together, and that's ok.

There are lots of ways you can get involved in Metadata2020 if you're interested. If you'd like to contribute to a project in your area of expertise (even just for a set period of time), you can email info@metadata2020.org for details. You can also help promote the initiative to the wider community through your organizations, word of mouth, and social media - [Twitter](#), [Facebook](#), [LinkedIn](#), and at metadata2020.org. Get involved and help us improve the quality of metadata for research!

Biography

Rachael has worked at Crossref in various roles since 2012. She is a participant in the Metadata2020 Researcher Communications group and advocate of richer metadata. She can be contacted at: [@rachaellammey](mailto:rlammey@crossref.org). This article is based on Rachael's presentation at the CILIP Cataloguing and Indexing Group (CIG) conference in Edinburgh, September 2018.

ORCID: <https://orcid.org/0000-0001-5800-1434>

Introduction

The IAEA Library was established in 1957 to provide scientific and technical information services to IAEA staff, Member States' permanent Missions in Vienna, official IAEA meeting participants, and affiliated researchers from around the world. It is a special library focused on the peaceful applications of nuclear energy.

The catalogued collection is relatively small – with only around 108,000 bib records. In 2015 it was decided that the library should adopt a new Library Management System. At the time there was only one library staff member involved with cataloguing and metadata related issues. In preparing the data for migration I worked closely with the Library's IT support professionals and the transition to the new LMS was only possible due to that collaboration.

Why a new LMS?

To paraphrase the great John Donne, we needed a system where *“no module is an island, entire of itself, every module is a piece of the system, a part of the main”*

1. Age

The existing system had been customised to the requirements of the library as they stood over 15 years ago. While upgrades and further customisations would have been possible they were not implemented and so the system had truly reached a point of not being functional for use in the 21st century.

In addition, knowledge transfer was lacking – there had been a lot of staff turnover in 15 years and no remaining experts meant that there were parts of the system which were not used and which no one knew how to use.

2. Lack of integration

Much like Fight Club, what happened in one module stayed in one module. For instance, bibliographic records created in the 'Acquisitions module' did not appear in the 'Cataloguing module' and were therefore created twice. Changes made to a record in one module were not reflected across all modules. Additionally, if you deleted a bibliographic record it did not delete any related Order or Holdings records. This alone caused some significant issues for the data migration.

3. Lack of user friendliness

On the staff side of the system, searches could not be refined once you had the results list. So, for example, a search for “radioactive waste management” would produce 454 'hits' and these could not be narrowed down via date of publication, author, call number, format or any other facet. Prior to conducting the search, the staff member could input a complicated search string such as this one designed to find monographic items catalogued as proceedings and published in 2013:

```
ITYPE=Proceedings AND MTYPE="monographic item" AND BDATE=2013
```

1. The original presentation given at CIG18 was co-authored by Jaime Garcia Llopis, Systems Support professional with the IAEA. This paper summarises the systems side of the migration but does not focus on it.

4. Not RDA compatible

We could not add new fields to the templates. The existing system did not allow staff to work with the leader field and did not use control fields so changes required to make those MARC21 fields RDA compatible could also not be made.

The lack of RDA compatibility was an increasing issue for us as we downloaded records from OCLC and were required to “revert” RDA records to AACR2 to make them compatible with our system.

5. Batch processes

Library staff were unable to batch edit or make global updates to records. All such changes needed to be referred to IT support. Even the simplest thing, such as changing a monograph record to a serial record had to be referred to IT. Anything along the lines of “global updating” also had to be referred to IT and often, because of the nature of the system, global updates could not be made at all.

A related issue was the lack of an effective URL checker in the existing system which meant that finding non-working URLs was almost impossible and solely done via manual checking of records following reports from users.

6. No formatting standard

The cataloguing templates did not fully conform to MARC21 nor to any other formatting standard. The best that could be said is that the existing system used a formatting standard which was “MARC adjacent”.

Issues in Preparing for the Data Migration

While massaging our records into MARC21 format was by far the greatest issue we faced, it was not the only one. Other issues included:

1. Lack of key staff familiarity with MARC21

No one involved in the data migration had learned MARC21 or used it in their day to day work.

2. Insufficient weeding of the collection prior to migration

Weeding was not possible prior to the data migration because of limited time, limited staff and issues related to taking inventory in the old system

3. Record duplication (copy control issues)

Individual items could not be suppressed which led to a lot of extra bib records – we needed to determine how to attach these item records to one bib record. Record duplication also occurred due to the way in which serial and order records were created – often leading to requests for new bib records which ended up being duplicates.

4. Mapping of local record fields

We had to make decisions about where our ‘local’ information would migrate to, whether all of it needed to be migrated and how to ensure that certain information remained suppressed from the OPAC.

5. Record ID number issues

We had a set of over 5,000 records which had a different sort of record ID number and this was only discovered after the first test load.

6. Fields that served multiple purposes

Subjects went into one set of template boxes with no way to differentiate between LC subjects and subjects added from different thesauri or, for instance, to show which subjects were MARC 610 vs 650. There was also no way to add field or subfield tags.

ISBN, ISSN, technical report numbers and standard numbers were all entered into the one set of boxes marked “report number”.

All 76X-78X and 580 notes were entered in one box.

7. Issues relating to item records

The item records in the old system were somehow part of the bibliographic record – you could not transfer items between records - you had to delete the item (basically the barcode) from one bib record and then enter it again on the preferred bib record.

There was also an issue that our “Collections” were our “Locations” – we have Circulation, Reference, Journal, Technical Reports, Standards and Online Collections and these equate to Locations. This caused a problem for the data migration – not on our side but from the side of the new LMS. The concept that the Collection and the Location were the same was somehow very confusing and caused a lot of problems with scoping for catalogue searches.

Essentially, we had to figure out from our limited “item information” how to create separate item records and attach them to the new bib records while keeping location, collection and item type correct.

Data preparation

The main issue we had (with our lack of MARC familiarity) was determining how to map the fields from our existing system to MARC21 – while much of this was straightforward (title goes to 245, author goes to 100) there were issues relating to fields where the existing system lumped everything together – so ISBN, ISSN, technical report numbers, standard numbers, government document numbers were all entered into one field in STAR and finding the ways to identify which was which for the data conversion could be very complex. In addition, we found many other issues such as serial records with no bib record. We had to decide how to handle staff papers (which were a customised record format, difficult to map properly to MARC21 and missing key fields like publisher etc.).

We had problems with the subject field (no indication of what thesaurus was used, no breakdown of subfields, and, finally, we had to identify all the different ways that one concept (such as ‘frequency’ or ‘language’) could be expressed

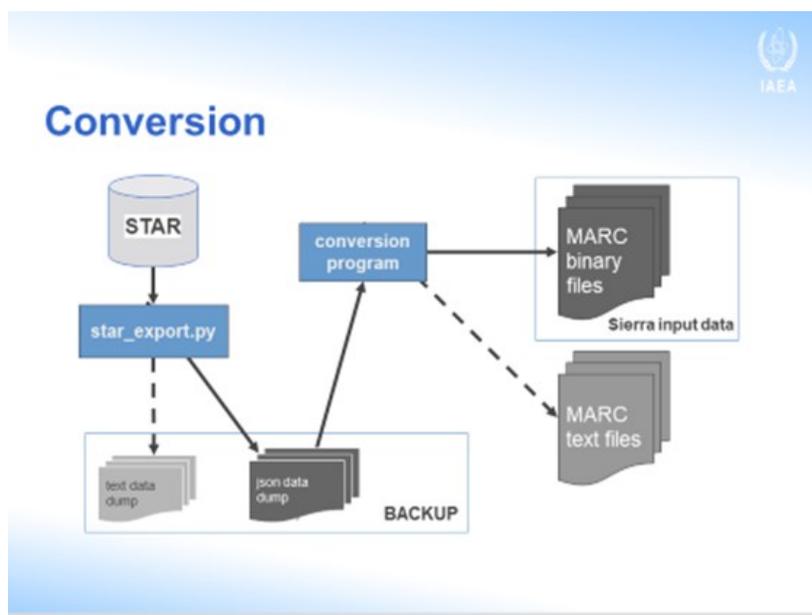
A minimal amount of data clean-up was done prior to the migration – obsolete locations were deleted, records were de-duplicated and small batches of records were manually fixed for issues such as “no language” or those missing a location.

Data extraction, conversion and testing

The tool chosen to manipulate the data was Python, specifically the PyMARC module of the Python library. While PyMARC is often used in conjunction with MarcEdit this was not the case for us – we did not use MarcEdit at all.

Once the fields from the existing system had been mapped to MARC21 then an API (from the existing system) was used to “dump” the data.

A script (star_export.py) was developed which provided output in 2 parallel formats – one was text to be used for validation and the second was json and the json file was used for the conversion.



Testing

We performed two test loads before the final migration. These tests highlighted several issues including:

Punctuation - some adjustments needed to be made to insert correct punctuation in the fields. This was partly due to our lack of familiarity with what was required in the data transfer – we had assumed that standard punctuation would be inserted during the migration without us necessarily having to script that.

Formats – it would seem simple but knowing how to know something is a technical report versus a standard versus a journal issue proved difficult and the test loads showed us where we needed to do more refining.

Nonfiling characters – particularly in languages other than English.

Subjects – despite our best attempts in the end we had to send all subjects as 650, second indicator zero and all subfields as |x

Field display– in our existing system the 500 field was always suppressed so had been used for a variety of notes – some which should not be displayed. When we saw that the 500 would be displayed in the new system (and decided that that would, going forward, be acceptable, we then had to change our local notes to 590). We also had a local subject field which needed to be displayed. The default for the new system included displaying 336, 337 and 338 which we only realized after the test load. These fields were suppressed from display.

Rejected records (eg. 505 field) – the test load showed us which records needed more work before the migration – the main instance of this was where our table of contents fields included more than 10,000 characters.

Diacritics – many didn't transfer correctly in the first test load

Post migration

Since the go-live in September 2016 there have been over 480,000 modifications made to records (mainly via global update processes). The major issues addressed included:

- Removal of duplicated fields
- Amending subject headings to ensure accurate subject subdivisions
- Inserting punctuation into fields
- Adding tags for title field (eg. |b and |h)
- Fixing item types
- Fixing locations
- Editing local subject field
- Adding missing language codes
- Removing duplicate 1xx fields
- Inserting punctuation in 711 fields
- Changing bib level from serial to monograph for journal special issues (which had been migrated as serials because of the presence of an ISSN)

Lessons learned

Negotiate a generous lead time – we initially had 7 months to complete this project from start to Go-live. We pushed the time frame out by one month but I think that while we succeeded it would have been better to have more time.

If possible, **weed beforehand** - the less records you must worry about migrating the better. We are now in the middle of a largescale weeding project and much of the material being discarded has bibliographic records which caused us the most problems during the migration.

The **project team** is the most important part of the actual data migration. It needs to consist of people with both IT and library side knowledge and those people need to be able to work together and support each other through what will be a very intense and at times supremely frustrating experience.

Work with your strengths – we had someone who knew Python and no one who had used other tools such as MarcEdit so we chose to work with what we knew.

Be prepared to make mistakes – and be prepared to be OK with that.

Make **strategic decisions** on where to focus your effort – we knew that in a lot of instances it would be easier to fix the data after the migration than before so we focused on those things which had to be correct on Day 1 and kept track of those things which would be corrected later.

Be prepared to be working for a long time after the migration to “fix” the data.

Make sure you still have **access to your old LMS** to check problems and resolve issues.

Further reading

Bannerjee, K., and Parks, B., (eds.), 2017. Migrating library data : a practical manual. American Library Association.

Wang, Z., 2009. Integrated library system (ILS) challenges and opportunities: a survey of U.S. academic libraries with migration projects. *The Journal of Academic Librarianship*, 35(3), pp 207-220.

Biography

Penny Doulgeris holds an MLS from Charles Sturt University in Australia and is currently the Metadata Librarian for the International Atomic Energy Agency (IAEA) in Vienna. Her work focuses on projects to enhance the quality of the library's metadata to facilitate search, discovery and access to the library's resources.

P.Doulgeris@iaea.org

ORCID ID: [0000-0002-2942-8049](https://orcid.org/0000-0002-2942-8049)

You might be somewhat wary of a paper where the title has to be explained, but *The LSE, the Blogs and the Metadata* is an allusion to the second chronicle of Narnia, *The Lion, The Witch and The Wardrobe*. Inspired by the wardrobe's power to let the children into another world, this paper looks at the power of metadata in accessing LSE's blog content. Just as the children tumbling out of the wardrobe, when they returned home at the end of C. S. Lewis' second chronicle, was only the beginning of their adventures in Narnia in the remaining chronicles, the project this paper describes is only the beginning of our blog adventures at LSE. It considers the key role of the Library, and specifically metadata, in supporting institutional goals and the wider work and outreach of the university.

In May 2016 the Library had the opportunity to bid for some Higher Education Innovation Funding (HEIF). This funding is given to support knowledge exchange between universities and the wider world which result in economic and social benefit to the UK.¹ Such a focus on social benefit is an integral part of LSE, which was established in 1895 for '*the betterment of society*'. The phrase we use 123 years later for sharing research in order to make a difference in how problems are understood and addressed around the world is '*Knowledge Exchange and Impact*'. Just one of the many ways academics are encouraged to do that is by blogging about their work to bring it into a sphere for wider academic communication.² A 2017 study of 45,000 academics and scientists carried out by the School of Informatics and Computing at Indiana University found that content from blogs.lse.ac.uk made the top 15 list of academic content that such groups share on social media. For political scientists, content from LSE blogs was the number one source of academic content they shared on Twitter. For sociologists it was number two and for economists it was number five.³

We know that LSE-generated blog content is significant for the institution, and yet by its nature it is entirely web-based content, which OCLC's 2018 report, *Descriptive Metadata for Web Archiving* describes as 'volatile', stating that '*if not preserved on a timely basis a significant percentage of web content simply ceases to exist*' which means it is '*imperative that we preserve web content on a timely basis if we are to maintain the integrity and continuity of the historical, cultural and scholarly records*.'⁴ This awareness had already informed Library Strategy and our 2015-2020 strategy includes a specific action to '*secure the collection and preservation of the complete intellectual output of the School*.'⁵

1. Higher Education Funding Council, *Knowledge Exchange Funding*, <http://www.hefce.ac.uk/ke/heif/> accessed 27/09/18

2. London School of Economics and Political Science, *An Introduction to Knowledge Exchange and Impact at LSE*, <https://info.lse.ac.uk/staff/services/knowledge-exchange-and-impact/Assets/Documents/PDF/Toolkit/1-Introduction-to-KEI-at-LSE.pdf> accessed 27/09/18

3. Ke, Qing, Ahn, Yong-Yeol, and Sugimoto, Cassidy R., *A systematic identification and analysis of scientists on Twitter*, PLoS ONE 12 (4): e0175368, 2017, <https://doi.org/10.1371/journal.pone.0175368> accessed 27/09/18

4. Dooley, Jackie, and Bowers, Kate *Descriptive Metadata for Web Archiving: Recommendations of the OCLC Research Library Partnership Web Archiving Metadata Working Group*, Dublin, Ohio, 2018, pg 5 <https://www.oclc.org/content/dam/research/publications/2018/oclcresearch-wam-recommendations-a4.pdf> accessed 27/09/18

5. LSE Library, *LSE Library Strategy*, <http://www.lse.ac.uk/library/about/lse-library-strategy> accessed 27/09/2018

Furthermore the School's 2015 Knowledge Exchange and Impact Strategy includes a specific action for the Library to archive all official blog output. Having an efficient, systematic and comprehensive approach will protect this content from the risk of loss in the future. By archiving the content in LSERO we ensure that its metadata feeds through to our Library Search platform, so that it is discoverable and accessible alongside library content.

The volume of blog content at LSE, however, makes all this no mean feat. We produce 61 blogs, as well as continuing to provide access to various closed blogs, and also contribute to a number of partner blogs.⁶ On closer inspection 56 of these were deemed to be 'official' output, with the internally focused community blogs, which do not include research, being outside the scope of the project. There were already some links with LSE Research Online (LSERO), the School's Institutional Repository, in cases where an author had specifically requested that their blog content was added to the repository. In 2016 this self-selecting content was an average of 63 blog posts a month, which is about a fifth of the School's monthly blog output. Our HEIF bid focused on extending and developing existing activity to retrospectively archive all official blog outputs and establishing automated procedures to make complete archiving feasible and sustainable in the future. If time allowed we also hoped to review the potential for adding DOIs to blog posts to improve discoverability and to offer the ability to measure impact through tools such as Altmetric.

We had initially anticipated hiring a project manager to bring niche technical skills to the project, but it became apparent that for a short term project the use of internal staff, with an existing knowledge of institutional practices and procedures, would be more beneficial, even where that meant factoring in some development time. However there were time restrictions attached to the funding, so it was not possible to upskill in every area of the project and we determined that the investigation of DOIs would need to be part of future plans. Our focus for this project was on retrospectively archiving the content, which meant creating PDFs, saving them in backed-up Library server space, loading them to LSERO and creating associated metadata, and establishing automated procedures to make the process sustainable. This ensured that we would have practical deliverables. Considerable initial investigative and planning work was required, including a workflows audit to check that the amount of time we thought it took to create blog records on LSERO was still correct. This allowed me to 'number-crunch' on the basis of one member of staff being able to add 200 records to LSERO in a week. However, even armed with this information, estimating project workflows proved much more difficult than I had hoped. The top ten blogs, in particular, frequently re-post each other's content, so some posts were duplicates which did not need adding but, as cross posting is only indicated at the bottom of a post, each one required manual checking. In the end we worked with two estimates; the maximum being based on the total number of posts on each of the blogs (though, for various reasons, even this was not always easy to determine) and the minimum being based on an educated guess at the number of unique posts using a sample of two months from each blog. Adding all this up, it was apparent that we had a minimum of 5800 posts to add, and a maximum of just over 20,000, which is not the kind of mathematical discrepancy you want, particularly when you are responsible for the project! We recruited three temporary staff to work on creating PDFs and the associated metadata, so if they all worked at the anticipated rate (an unknown at this stage) we would create 600 records a week, and adding all content would take us somewhere between ten weeks and 33 weeks – but our master finance spreadsheet told us we could fund three temps for a maximum of 16 weeks. This was a particularly important time for managing expectations, and we focused our aim and communications on archiving official content from the top ten blogs. Stakeholder engagement is, of course, an essential part of project management and we liaised with the School's social media manager who invited us to meet the top ten blog editors and outline the project. They were encouragingly enthusiastic, and it quickly became evident that there was a real appetite in the School for this work to take place.

6. London School of Political and Economic Science, *LSE Blogs*, <http://blogs.lse.ac.uk/> accessed 27/09/2018

In an ideal world, we would have worked on establishing the automated processes for adding content before we employed the temps so that more could be achieved within the short time available, but due to the funding and timing constraints of the project we needed to work on both aspects simultaneously. The first step was to automate the creation of PDFs of each blog post. Our colleagues in IT were able to do this for us and they now put these into a shared file which we can pick up every month. For the metadata we have been able to semi-automate its creation through the use of BibTex files which can be imported along with the PDFs. This automates the process as far as is currently possible with the blog templates being used by the School. Some metadata is consistent across all blog posts, such as publisher, copyright or item type, and can be put into a template. Metadata automated from the live posts, such as date, blog name, post title, base and blog URLs can then be added to this. Despite creator, subject and description being three of the five most commonly used Dublin Core elements for archiving web content,⁷ we have been unable to automate the metadata for those fields. In our blog templates the 'author' is the person who published the post, who tends to be the editor of the blog, rather than the author of the post content. The author of the content is only identified in the body of the text rather than as a separate field. We receive the blog abstract as part of the automated metadata feed, but due to a lack of consistency in the way it is formatted across multiple blogs, it often truncates part way through, and so requires manual checking. The subject tags on the blogs are not the same as the scheme we use in LSERO but, even if they were, we understand that we would not be able to auto-populate the subject trees in Eprints. So we are limited to partial templates, but this semi-automation is still better than the entirely manual process we would otherwise have, and it has enabled us to shave three minutes off the process for each record (from ten to seven minutes).

After 16 weeks of dedicated work by our temps, our project expectations were exceeded with the addition of 11,665 blog posts to LSERO, the total unique content from all of LSE's 56 official blogs. Having also completed the automation aspect of the project there was now some further 'number-crunching' to do in order to investigate how sustainable it was to archive blogs in the future.

$$\begin{array}{r}
 \text{Average number of blog posts for each blog per month} \\
 \times \\
 \text{New time to create a record on LSERO} \\
 + \\
 \text{Average number of re-posts per blog per month and time taken to eliminate them from the process} \\
 = \\
 \text{12 hours work a week to keep up with archiving all official blog output}
 \end{array}$$

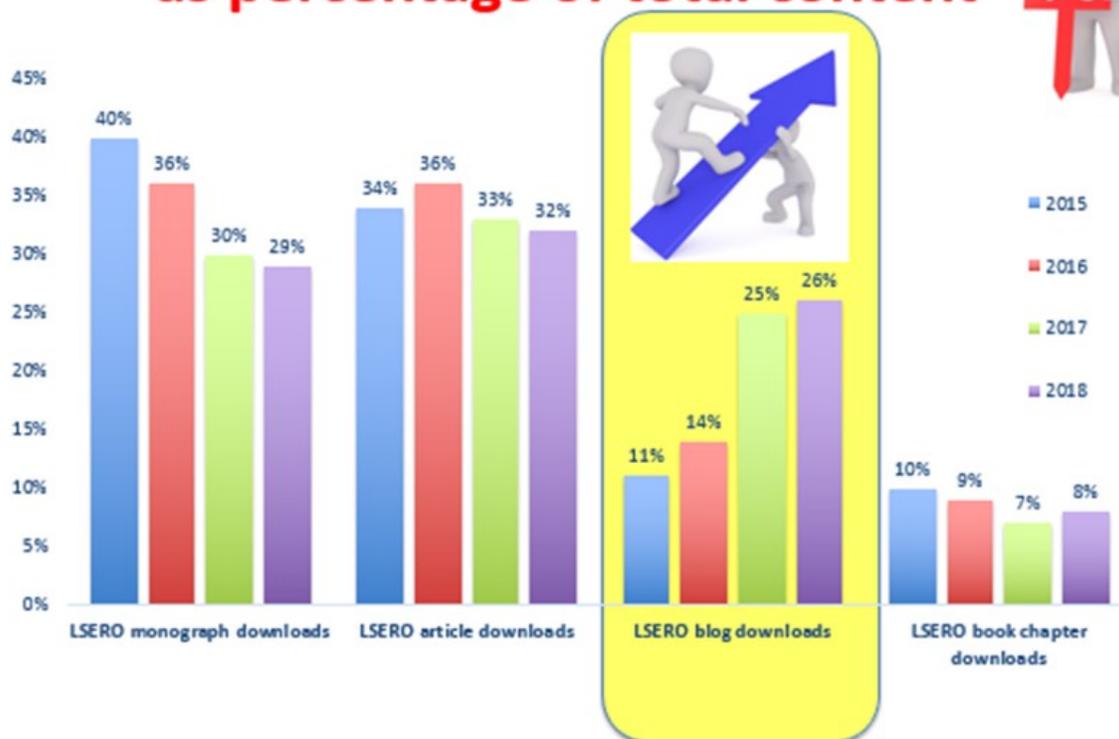
I do not imagine anyone reading this article is operating in a team which could simply absorb that amount of extra work, nor that any reader could magically produce the finances to increase resourcing in their team by nearly 0.4 FTE. It was time to return to the 'managing expectations' aspect of the project, which had made clear that the top ten blogs were the priority. Digging into the figures a bit more it became clear that in theory archiving content from just those top ten blogs would take an average of six hours a week, and that all the other blogs put together made up the remaining six hours a week. Our team does not, of course, have six hours a week to spare. Nevertheless, six hours is significantly less than 12, and we wanted to be offering something ongoing from the project, so we took the risk of saying that we would trial absorbing those six hours into team workflows, but that beyond that we would require additional resourcing. In practice, looking at the statistics, this work has taken six and a half hours a week, but the self-selecting blog content that we had been dealing with for authors before this project began took two and a half hours a week, so we have actually only an additional four hours to add to usual workloads.

7. Dooley and Bowers, *Descriptive Metadata for Web Archiving*, pg 12 <https://www.oclc.org/content/dam/research/publications/2018/oclcresearch-wam-recommendations-a4.pdf> accessed 27/09/18

We have four members of staff to absorb this, so effectively each of them had to find an extra hour a week for the blog content, which suddenly sounded more hopeful. Each of those four staff takes responsibility for one or more named blogs, based on the average output for each blog, so that they are each receiving a similar amount of work. This worked excellently for four months, but then a prolonged staffing vacancy coinciding with a further new project significantly reduced our capacity and we had to put the blog work on hold for a period of months, returning to it about six months later when resources allowed.

As we closed the project, we celebrated meeting a specific goal given to the Library by the School's Knowledge Exchange and Impact strategy to archive all official blog output up to July 2017, ensuring that this key content, which falls outside the more traditional publishing channels, is protected from risk of future loss. We were also able contribute to the specific Library Strategy action to secure the collection and preservation of the complete intellectual output of the School by adding all official blog output to date. New workflows mean that we can keep up with work to secure the content of the top ten blogs each month. Content outside the top ten blogs has not been collected since the end of the project, but we are now in a much better position to do this retrospectively, should resourcing become available. Given that the blog content is still live on the LSE blogs themselves, it has been important to consider whether our work has had an impact on the visibility and discovery of this content. As might be expected, accessing blog content directly from the blogs themselves is currently its primary access point. The blogs have more than 70,000 unique users each week, and the most popular post has had 470,000 views. We cannot rival that in LSERO, but we can point to an increase of activity in accessing blog content on LSERO as illustrated in the graphic below. (Accounts for 95% of downloads with the others being conference papers, AV etc.).

LSERO item type downloads as percentage of total content





Blogs are still third place in terms of most viewed content type, but they have had a significant percentage increase compared to other content types.

The success of the project led to a number of positive responses from blog editors and academics, raising the profile of the Library as a partner in disseminating LSE's research outputs.

I mentioned earlier that there had been timing constraints on our project, restricting some of the work we had hoped to do, with the result that when resourcing allows there are areas we would like to work on as a second phase of this project. OCLC's *Descriptive Metadata for Web Archiving* report addresses various issues around the capture and creation of metadata for web-based content, so we may wish to review the metadata applied to our live blog posts and consider how that could be purposed to improve discoverability. This would require liaison work outside the Library, discussing the School's blog template, and agreeing what level of consistency could be established across the blogs to facilitate more efficient application of, and subsequent capture of, descriptive metadata. We would benefit from including LSE's Website Improvement Team in this review as we think about how best to surface blogs alongside other LSE content. Our Digital Library team have developed a DOI minting service, so there are opportunities to investigate adding DOIs to blog content, both to give increased stability to content in terms of permanence and discovery, and to consider how we might then be able to measure the impact of this content via tools such as Altmetric. Finally, we could consider auto-classification for this content. *Descriptive Metadata for Web Archiving* indicates that we are not alone in struggling with the need for 'scalable descriptive metadata practices that take into account the extremely limited human resources available'.⁸ The subject classification we use for blog content, coupled with the fact that lack of staff time is our highest barrier to metadata creation for that content, means that blog content could benefit from our experimenting in this area. It would require more investigation to determine feasibility, but we do need to consider options that allow us to use metadata expertise in scaling up the blog archiving operation with our limited resources... so do watch this space, because I hope that what I have written above is just the beginning of our adventures!

Biography

Helen is the Metadata Manager at LSE Library where her team have responsibility for print, electronic and institutional repository metadata, with a strategic focus on metadata management to support discovery. Helen has previously worked for the Institute of Directors, and The London Library, and was on the CIG committee from 2009-2016.
H.K.Williams@lse.ac.uk
[@HelsKRW](#)

Images CC0 1.0 Universal (CC0 1.0) [3dman-eu: Pixabay](#)

8. Ibid, pg 11

1. Introduction

Aston University had no metadata specialist for around 20 years so when I started in 2016 there were a number of challenges to manage. To name a few, authority control, classification and metadata quality all needed reviewing. I needed a cost effective approach to tackling these and as the only member of staff with professional expertise in this area manually managing the data was not an option.

1.1 Background and history

In 1995 the library underwent a major restructure. This new structure meant that each professional member of staff would do a little of everything. Until November 2016 there was no professional metadata or cataloguing specialist in the library. In the meantime cataloguing was managed by one professional member of staff and a few paraprofessional staff who did all the cataloguing work. In 2008 the library went almost completely shelf ready for printed stock, and all e-resource records were vendor supplied.

I was appointed in November 2016 as a metadata professional and tasked with cleaning up the database. Some of the issues I discovered in the catalogue in no particular order were:

- A lack of authority control
- Missing fields
- Incorrect data in fields
- Local subject headings and name authorities (not related to LCSH)
- Local indexes
- No Library of Congress Subject Headings (LCSH)
- Hybrid e-book and print records
- Split multi-volume works

This is by no means a fully comprehensive list of issues but demonstrates some of the major problems. My approach to improvements was to prioritise things that both improved workflows and efficiency or, less tangibly, improved access and discoverability of our collections. My first efforts were directed to improved workflows for the Information Assistants, to save time in stock through put thus enabling them to have more time for other jobs.

1.2 MarcEdit

MarcEdit was the tool I opted to use for each of the metadata projects described in this paper. It was first created in 1999 by Terry Reese for a data clean-up project at Oregon State University. It is now offered to the library world for free and includes a vast array of functionality within it. I will describe in some detail how I utilised MarcEdit in four separate projects aimed to improve the library metadata. These projects were:

- Authority control
- Module codes updates
- Reclassification
- Metadata enhancement of legacy records

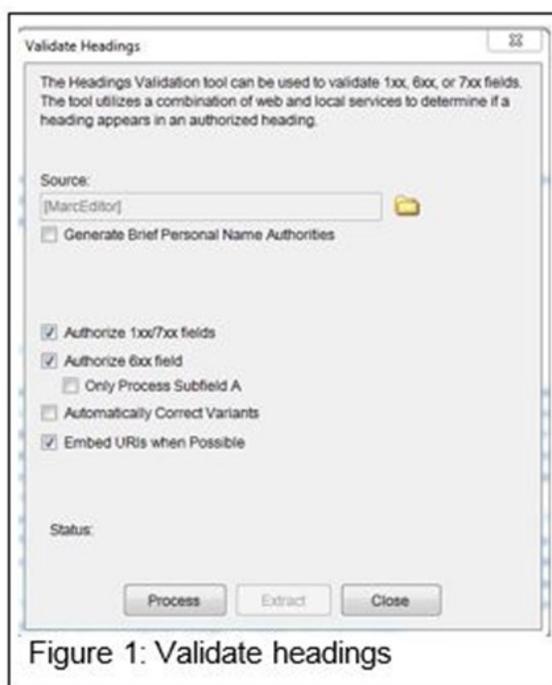
2. Authority control project

One of the first challenges I faced on starting at Aston was the lack of systematic authority control within the database. There were two authority indexes present, one for subjects and one for names. The subject index contained contentious headings and did not match any international standard, and the name authority index was basic and again conformed to no international standard. Additionally, these indexes were not updated either systematically or ad hoc and thus only covered material that was 20 or more years old. Upon review of the indexes I deleted all the local authority records in favour of implementing NACO headings for names and Library of Congress Subject Headings (LCSH) for subjects. Records in scope for the project were all MARC records with the exception of unpurchased PDA records.

I will describe in some depth the processes and steps I undertook in this project as many of them are replicated in the other projects as well.

2.1 Data extraction and manipulation

The first step is to identify the records that need authority control, for instance all print books or all single purchase e-books. In Sierra I create a list of these records and export the MARC records either saving them locally or in a cloud storage system (my preferred option). Once I have the MARC records I use MarcBreaker to create a mnemonic MARC file (essentially a text version of the MARC record that is easily manipulated and readable by humans). Once in the MarcEditor I use the Validate Headings function and select to embed URIs (figure 1). All the headings that can be validated will have a URI inserted which I can now use (figure 2). The URI contains in it the Library of Congress Control Number (LCCN) and I can extract this and use the LCCN to download the authority records. I do this by searching MARCEdit for all instances of “\$0http://id.loc.gov/authorities/names/” and copying this into Notepad++ (figures 3-4).¹



1. The regular expressions shown here are the ones I routinely use in Notepad++, however, this is not the only way to do the same thing, nor is it necessarily the best way. This is a good opportunity to try things out to see what happens.


```

=100 1\${MacKay, Donald MacCrimmon,$d1922-1987.$0http://id.loc.gov/authorities/names/n79099145 Jump to Record #: 11
=100 1\${Smart, Ninian,$d1927-2001.$0http://id.loc.gov/authorities/names/n50012584 Jump to Record #: 12
=100 1\${Braudel, Fernand.$0http://id.loc.gov/authorities/names/n80014791$1http://viaf.org/viaf/9843788 Jump to Record #: 13
=700 1\${Ollard, Richard Lawrence.$0http://id.loc.gov/authorities/names/n79097352$1http://viaf.org/viaf/64027639 Jump to Record #: 13
=700 1\${Reynolds, Sian.$0http://id.loc.gov/authorities/names/n86079925 Jump to Record #: 13
=100 1\${Braudel, Fernand.$0http://id.loc.gov/authorities/names/n80014791 Jump to Record #: 14
=100 1\${Braudel, Fernand.$0http://id.loc.gov/authorities/names/n80014791 Jump to Record #: 15
=100 1\${Marwick, Arthur,$d1936-2006$0http://id.loc.gov/authorities/names/n50052475 Jump to Record #: 16
=100 1\${Braudel, Fernand.$0http://id.loc.gov/authorities/names/n80014791 Jump to Record #: 17
=700 1\${Reynolds, Sian.$0http://id.loc.gov/authorities/names/n86079925 Jump to Record #: 17
=100 1\${Macintyre, Michael$c(Writer on Asia)$0http://id.loc.gov/authorities/names/n80092814 Jump to Record #: 18
=100 1\${Brandt, Willy,$d1913-1992$0http://id.loc.gov/authorities/names/n80138336 Jump to Record #: 19

```

Figure 4: List of names with URIs.

Using some simple regular expressions I can make the LCCN searchable via z39.50 (figures 5-7).²

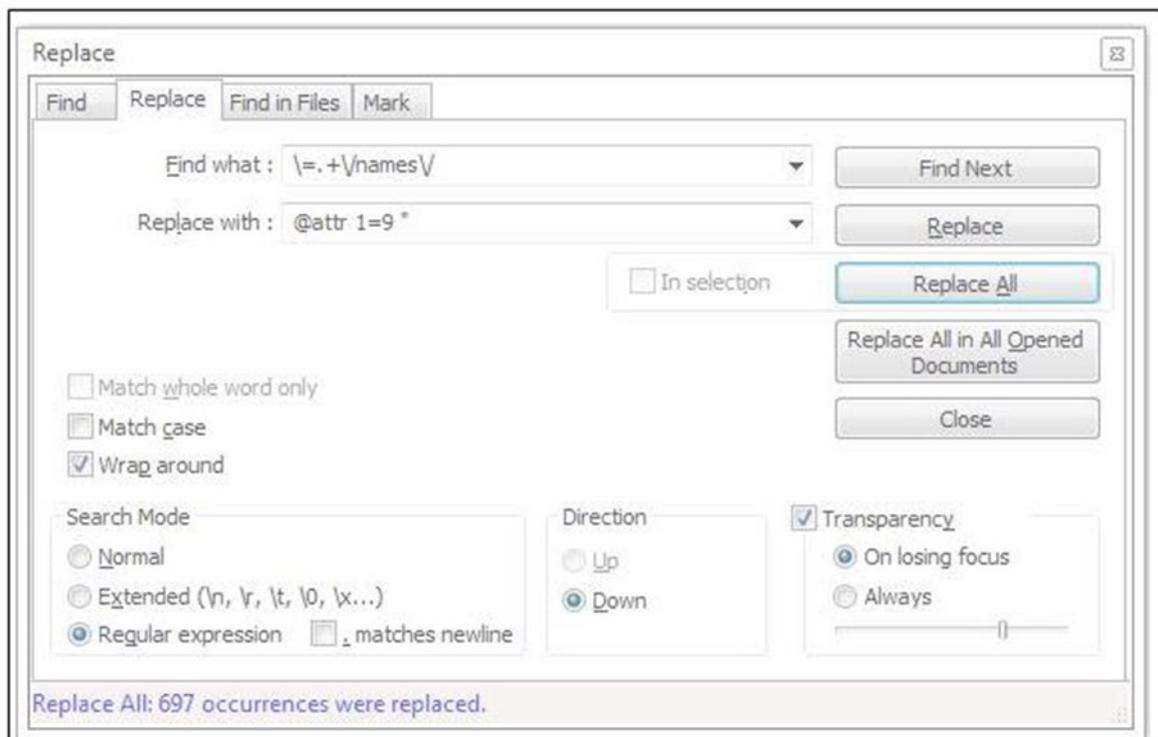


Figure 5: Regular expression to remove URI and insert z39.50 search criteria

2. A list of the attributes searchable via z39.50 can be found here: <https://www.loc.gov/z3950/lcserver.html>. In this case I search attribute 9 which in NAF is the LCCN in the 010 field.

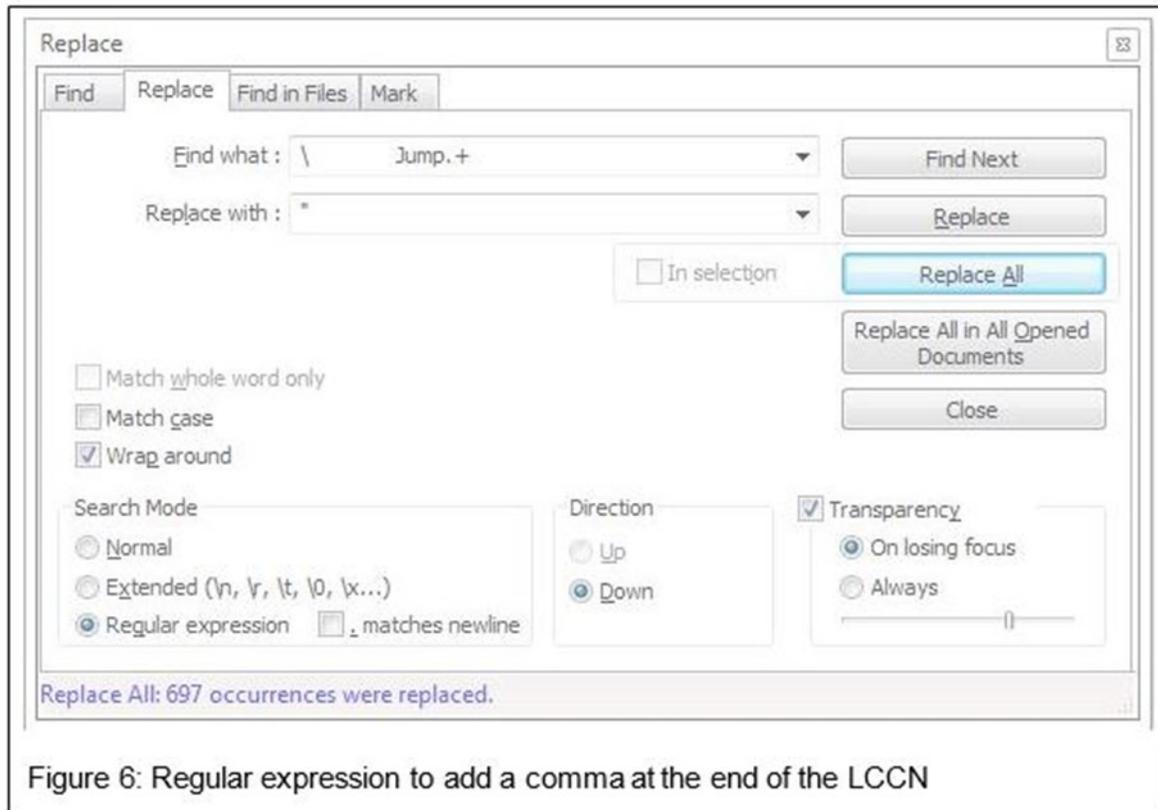


Figure 6: Regular expression to add a comma at the end of the LCCN

```

@attr 1=9 "n 50018377"
@attr 1=9 "n 81062380"
@attr 1=9 "n 79099145"
@attr 1=9 "n 50012584"
@attr 1=9 "n 80014791"
@attr 1=9 "n 79097352"
@attr 1=9 "n 86079925"
@attr 1=9 "n 80014791"
@attr 1=9 "n 80014791"
@attr 1=9 "n 50052475"
@attr 1=9 "n 80014791"
@attr 1=9 "n 86079925"
@attr 1=9 "n 80092814"
@attr 1=9 "n 80138336"

```

Figure 7: z39.50 searchable LCCNs

Using MarcEdit's z39.50 client in batch mode I search the Library of Congress Name Authority File (NAF) using the custom search option (figure 8). The resulting set of MARC authority records can now be loaded into Sierra.

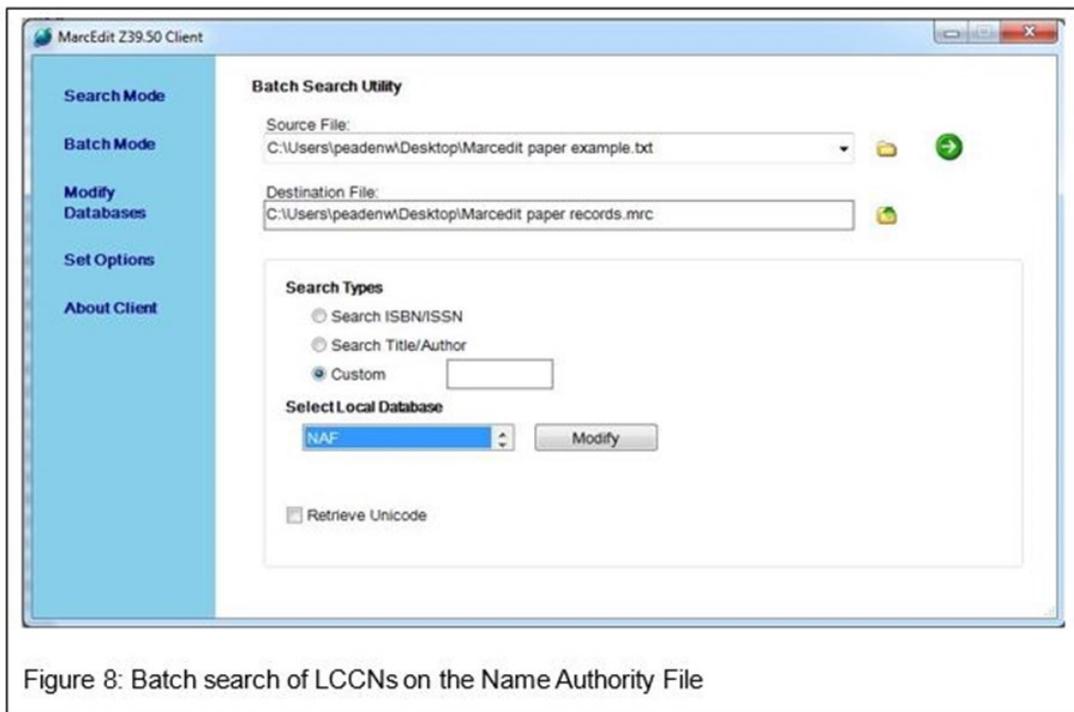


Figure 8: Batch search of LCCNs on the Name Authority File

3. Module codes in MARC records

At Aston we use Talis for reading list management and we put the module codes for the reading list into our MARC records. The reading list codes are indexed in Sierra and searchable in the catalogue by our users and staff use them internally for a variety of functions. In the past, the module codes were added and deleted manually by the Information Assistants. This process however, was both time consuming and prone to occasional errors.

3.1 Data from Talis to Sierra

I will now describe the process of taking the data from Talis and inputting it into Sierra. Figure 9 shows the data in Talis extracted into an Excel spreadsheet. It contains two key pieces of data, the module code and the local control number (LCN). With these I can create some “dummy” MARC records. Firstly, I will take those two fields and some others (author, title and ISBN) and separate them from rest of the spreadsheet (figure 10).

	A	C	D	F	G	H	I	K	N
1	Title	Item Li	Author	List Appearance	List Lin	Type	Import	ISBN13	LCN
2	The big book of dashboards: visualizing http://ast	Wexler; S	BN2292 - Dashboard	http://ast	Book		Recomm	978111928	b1775380
3	The Art of Taking Action: Lessons from J	http://ast	Krech; Gre	Wellbeing Collectio	http://ast	Book		9.78E+12	b1772486
4	Uni lifehacks: insights from teh UK's mc	http://ast	Surname; Wellbeing Collectio	http://ast	Book			9.78E+12	b1772485
5	Brock biology of microorganisms	http://ast	Madigan; BY2AM1 - Applied M	http://ast	Book		Further	9.78E+12	b1772484
6	Brock biology of microorganisms	http://ast	Madigan; PH2502 - Pharmaceu	http://ast	Book		Further	9.78E+12	b1772484
7	Brock biology of microorganisms	http://ast	Madigan; HS1MI1 (HS1MI1)	http://ast	Book		Further	9.78E+12	b1772484
8	Brock biology of microorganisms	http://ast	Madigan; HS1IS1 - Introductor	http://ast	Book		Further	9.78E+12	b1772484
9	Brock biology of microorganisms	http://ast	Madigan; BY2MI1 - Microbiolo	http://ast	Book		Further	9.78E+12	b1772484
10	Brock biology of microorganisms	http://ast	Madigan; BY1MI1 - Microbiolo	http://ast	Book			9.78E+12	b1772484

Figure 9: Talis data with two key pieces the code itself and the Local Control Number.

	A	B	C	D	E
1	Title	Author(s)	List Appea	ISBN13	LCN
2	The big bc	Wexler; S	BN2292 - I	978111928	b17753600
3	The Art of Krech; Gre	Wellbeing	978098242		b17724867
4	Uni lifeha	Surname; Wellbeing	978152721		b17724855
5	Brock biol	Madigan; BY2AM1 -	978129223		b17724843
6	Brock biol	Madigan; PH2502 - F	978129223		b17724843
7	Brock biol	Madigan; HS1MI1 (H	978129223		b17724843
8	Brock biol	Madigan; HS1IS1 - Ir	978129223		b17724843
9	Brock biol	Madigan; BY2MI1 - M	978129223		b17724843
10	Brock biol	Madigan; BY1MI1 - M	978129223		b17724843
11	Brock biol	Madigan; BI4050 (BI	978129223		b17724843
12	The psych	Svanberg; Wellbeing	978113820		b17724636
13	The psych	Rotenberg; Wellbeing	978113867		b17724624
14	The psych	Mair; Carc Wellbeing	978113865		b17724612
15	The psych	Ogden; Ja Wellbeing	978113850		b17724600
16	The psych	Taris; Too Wellbeing	978113820		b17724594
17	The psych	Wood; Ga Wellbeing	978113874		b17724582
18	The psych	Cotterill; S Wellbeing	978113821		b17724570
19	The psych	Cotterill; S Wellbeing	978113821		b17724570
20	The psych	Barker; M Wellbeing	978113867		b17724569
21	The psych	Gross; Ric Wellbeing	978113808		b17724557
22	The psych	Gross; Har Wellbeing	978113820		b17724545
23	Don't wait	Wood; Ga Wellbeing	978184112		b17724533
24	Unlock yo	Wood; Ga Wellbeing	978178028		b17724521
25	Flourish: e	Seligman; Wellbeing	978185788		b1772451x
26	Wireless \	Naja; Rola EE402B - R	978144195		b17724491
27	Wireless \	Naja; Rola EE4016 - M	978144195		b17724491

Figure 10: Separate spreadsheet with a few data points.

The additional data points are only for human readability of the dummy MARC records and to be able trace any errors on loading. The next step is to use MarcEdit's delimited text translator function (figure 11). This will translate data from a spreadsheet into mnemonic MARC records. Once these are compiled into MARC records I use a locally created load table in Sierra to input the module code. It matches on the LCN and imports only the 980 field. Once in Sierra I can transfer the data from the 980 field into the local index module code filed 900.

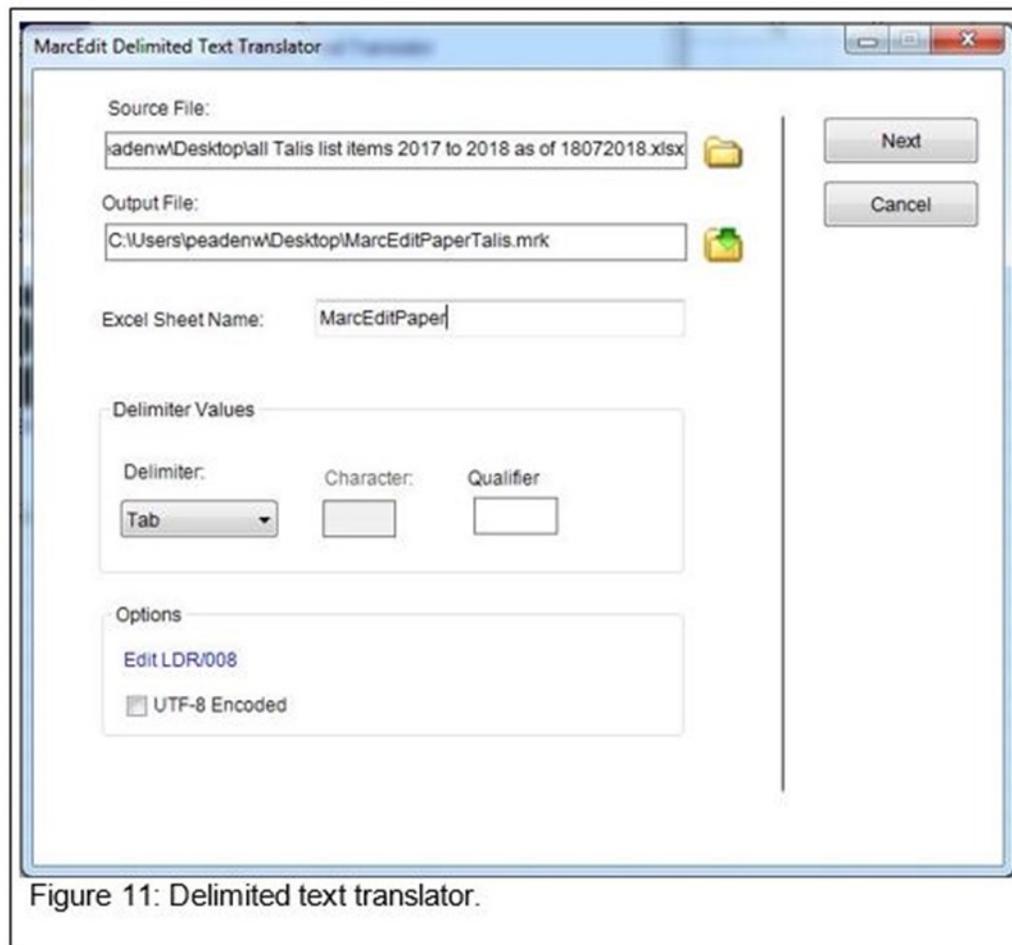


Figure 11: Delimited text translator.

4. Reclassification

Since 2008 almost all classification of stock was outsourced to our shelf ready vendors. These classifications rightly use the latest standard and generally follow the classification given by various agencies such as the Library of Congress. Over this time various split collections emerged, with books on a single subject being classed in two places. One major area was sociology (301-307 in Dewey) which had been classified to a local non-standard Dewey until 2008 and had various large scale “dump numbers” containing a variety subjects.

To reclassify this collection I first extract the MARC records in scope from Sierra. I use MarcEdit’s generate classification function (figure 15) which, instructs MarcEdit to look up the classification from OCLC’s Classify service and import it into a local MARC field 982. I use a similar load table to the module code project where it will insert the local 982 field based matching the LCN. Over the summer, the Information Assistants physically take the books off the shelves and copy the new classification into the 082 field. These books are then sent to be respined in the processing room.

The reclassification provided nearly 95% coverage of the records leaving a manageable amount to do manually. On the whole, the process is fast and easy to use and generally accurate and up to date. It is also possible to insert fast headings to the records at this time as well, though I opted not to do so. It is not perfect and some classification is not always correct, but these are relatively easy to deal with.

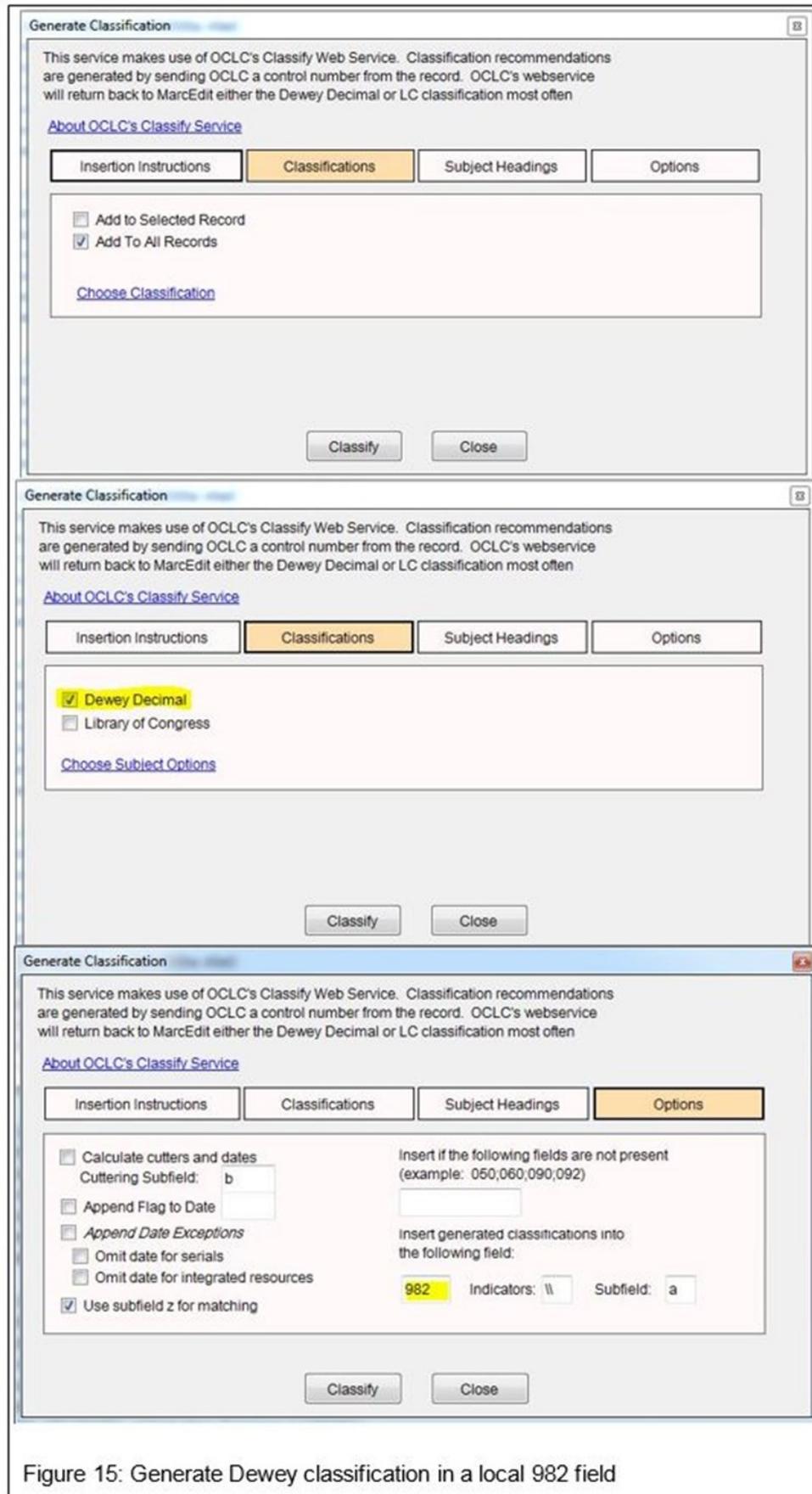
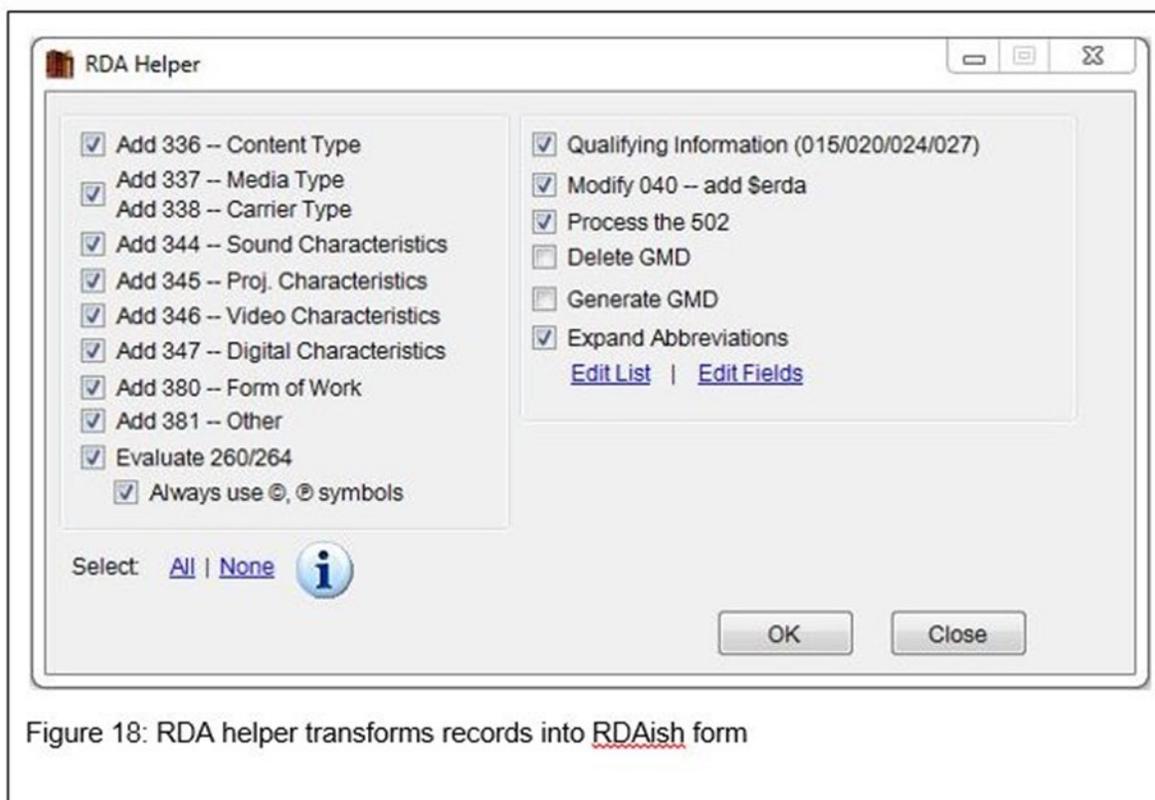


Figure 15: Generate Dewey classification in a local 982 field

Once the data is in a z39.50 query format I will batch search for the records in MarcEdit's z39.50 client. MarcEdit provides a report on how many records were found, which records brought back multiple results and which failed to find matches. This report is very useful for cleaning up the data and for forming a new query for a different database with the records that failed to match. I use MarcEdit's batch processes to make the records "RDAish", add linked data and clean up any invalid data using the various tools in MarcEdit that allow this (figures 18-21). Finally, I use Merge Records function in MarcEdit to merge the LCN of the record in Sierra (found in the 907 field) with the records obtained from elsewhere to serve as a matching point for loading into Sierra,⁵ and check the records from Sierra and the new records via a formula in excel (figures 22-23). Once all this is done I load the enhanced records into Sierra.



5. This is much more reliable than matching via a load table in Sierra and allows for any mismatches to be picked up prior to overwriting the data in Sierra.

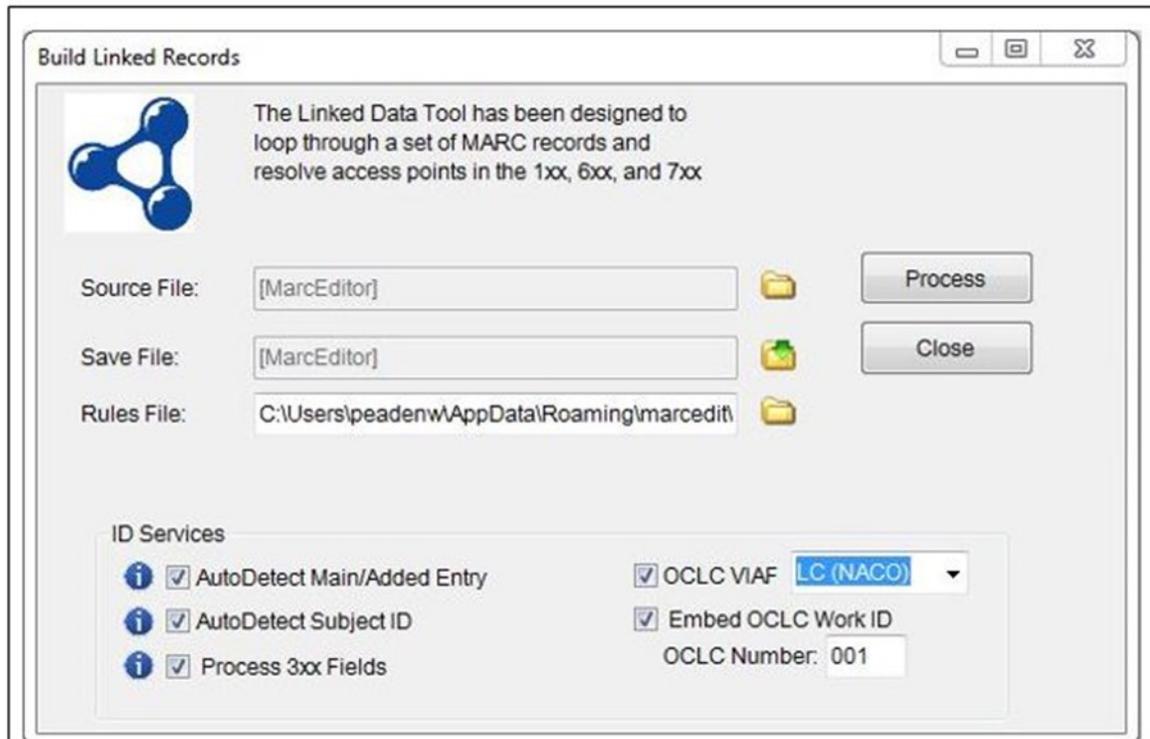


Figure 19: Add linked data elements. Does quite go as far a FrankenMarc but a nice feature.

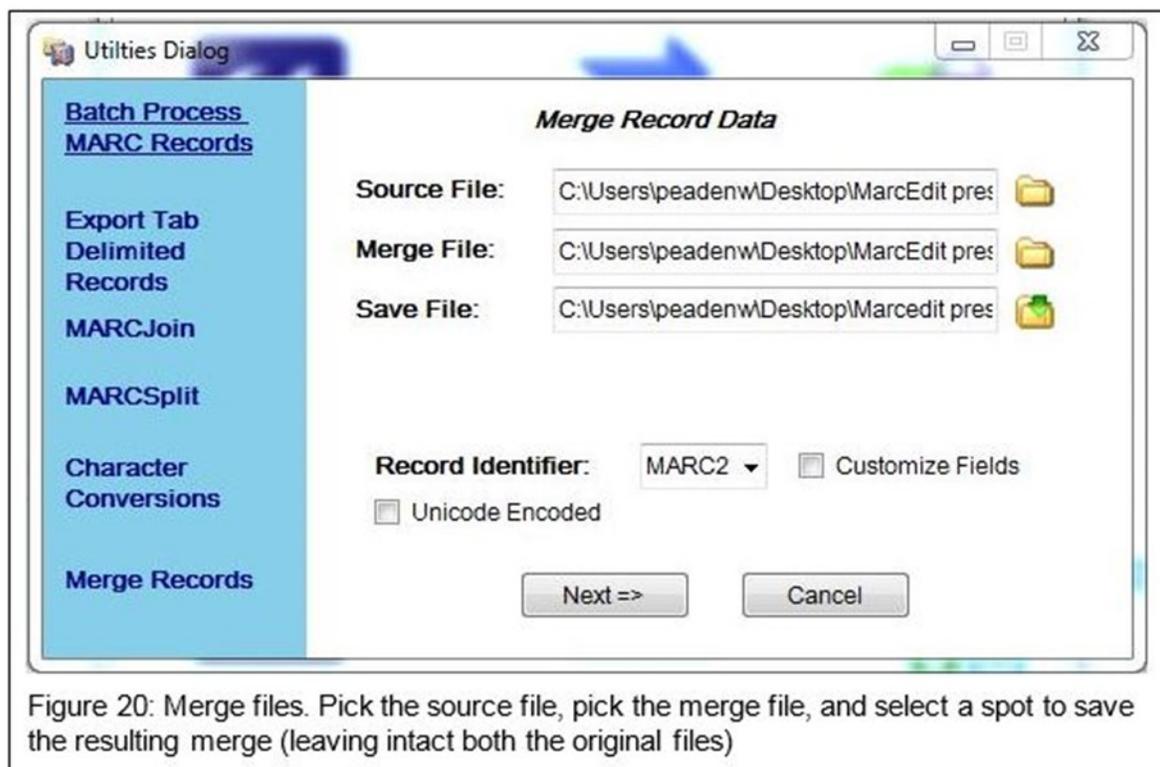


Figure 20: Merge files. Pick the source file, pick the merge file, and select a spot to save the resulting merge (leaving intact both the original files)

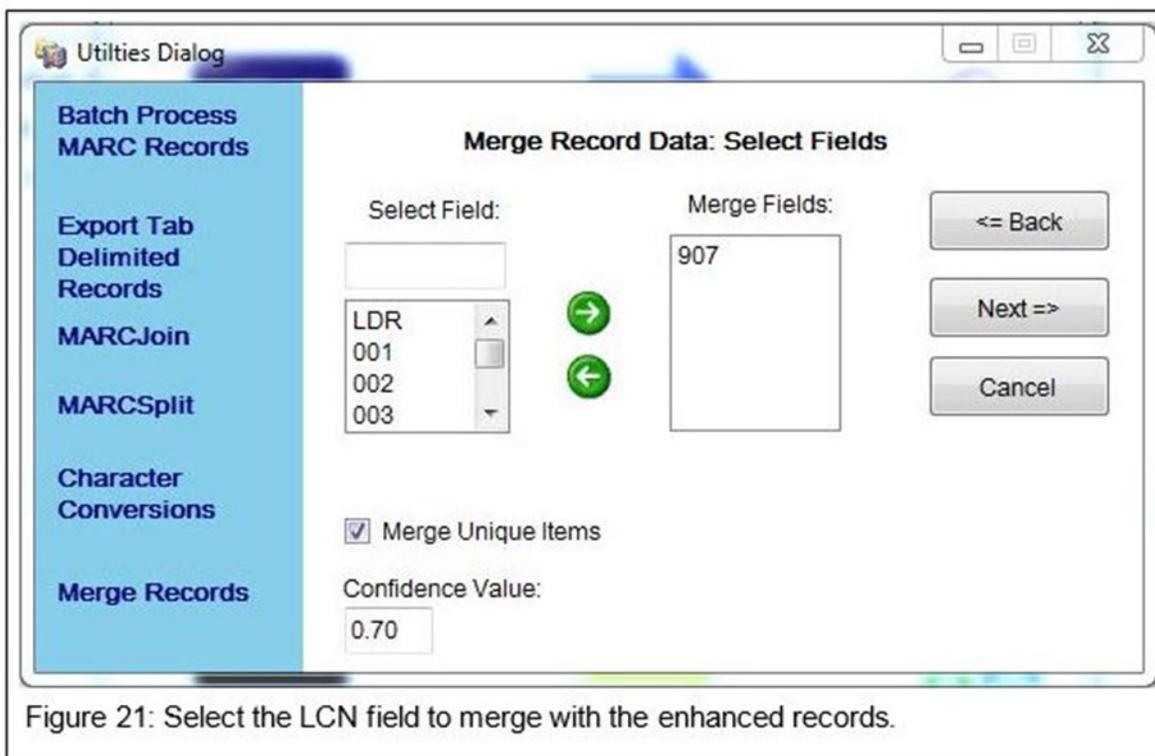


Figure 21: Select the LCN field to merge with the enhanced records.

```

=LDR 01353pam 2200301 | 4500
=001 799655
=008 821003s1974\|l|nyua\|l|b\|l|001\|leng\|
=010 \|a73022335
=020 \|a0030069416
=035 \|a0101-24360
=040 \|aDLC$cDLC
=050 04$aLB1025.2$b.G56
=100 1$aGivner, Abraham
=245 12$aA handbook of behavior modification for the classroom /$c[by] Abraham Givner [and] Paul S. Graubard
=260 11$aNew York,$bHolt, Rinehart and Winston$c[1974]
=300 \|axiii, 174 p. $bill. $c23 cm
=504 \|aBibliography: p. 169-170
=650 \0$aTeaching
=650 \0$aBehaviorism (Psychology)

```

Figure 22: Record before automatic enhancements (still better quality than the legacy record)

```

=LDR 01353pam 2200301 i 4500
=001 799655
=008 821003s1974\\inyual\\bl\\001\\leng\\
=010 \\$a73022335
=020 \\$a0030069416
=035 \\$a0101-24360
=040 \\$aDLC$erda$DLC$beng
=050 04$aLB1025.2$b.G56
=100 1\\$aGivner, Abraham$0http://id.loc.gov/authorities/names/no98078374$1http://viaf.org/viaf/32565513
=245 12$aA handbook of behavior modification for the classroom /$c[by] Abraham Givner [and] Paul S. Graubard
=264 1\\$aNew York :$bHolt, Rinehart and Winston,$c[1974]
=300 \\$axiii, 174 pages :$billustrations ;$c23 cm
=336 \\$atext$btxt$2rdacontent$0http://id.loc.gov/vocabulary/contentTypes/txt
=337 \\$aunmediated$bnc$2rdamedia$0http://id.loc.gov/vocabulary/mediaTypes/n
=338 \\$avolume$bnc$2rdacarrier$0http://id.loc.gov/vocabulary/carriers/nc
=380 \\$aBibliography
=504 \\$aBibliography: pages 169-170
=650 10$aTeaching$0http://id.loc.gov/authorities/subjects/sh85133052
=650 10$aBehaviorism (Psychology)$0http://id.loc.gov/authorities/subjects/sh85012905
=650 17$aTeaching$(OCoLC)1144565$2fast
=650 17$aBehaviorism (Psychology)$$(OCoLC)829955$2fast
=700 1\\$aGraubard, Paul S$1http://viaf.org/viaf/91590844
=758 \\$0http://worldcat.org/entity/work/id/1783002
=907 \\$a.b10021140$b18-10-18$c04-09-07

```

Figure 23: Record after automatic enhancements and with LCN.

6. Conclusion

I have described these processes in some detail and hopefully these examples are illustrative enough for duplication in other contexts. The main difficulties I found to overcome in the process is in the normalisation of data. For instance, in the authority control project I needed to get a list of LCCNs but it takes several steps and some cleanup of data before they can be useful. The main lesson is the importance of understanding what your data says and how it is structured so you can manipulate it to do what you need it to do. All of these processes developed as time went by and changed over the course of the projects. Trial and error and some Googling really pay off and making mistakes and figuring out why it happened help to deepen my understanding of the data as well as refine the processes. My advice to anyone interested in any of these projects is to try it out and see what happens.

Biography

Will worked at the University of Warwick as a Metadata Assistant and then as a Metadata Librarian from 2010-2016 and while there obtained his MSc in Library and Information Studies from Robert Gordon University. In 2016 he moved to Aston University to manage metadata and collections.
w.peaden@aston.ac.uk

Abstract

The University of Liverpool has a large collections of e-resources spread across many different platforms from many different publishers. To maintain accurate and fullest coverage of our collections on both Sierra Web OPAC and EBSCO Discovery system, we have established and enhanced our practices. We use the EBSCO holdings download as a primary source of data for particularly our journal collections and also collections of other materials where MARC record provision is none existent and alternative data sources relatively unwieldy to convert.

This paper details the development of this little exploited process from clunky and problematic to efficient and essential as a source of record provision for catalogue data to serve WebOPAC and EDS as well as SUNCAT, RLUK and Worldcat, with particular emphasis on load table exploitation and global update tagging to best manage data compatibility between the 2 systems.

This is an update on the presentation given during EIUG 2017, including details regarding improvements in load process and format management.

What's it all about?

When we were considering using the EBSCO MARC download service, we were in a situation where we were already maintaining the bulk of our collection in MARC format on our LMS. However, the quality of the records varied considerably. For the majority of our ebook collections, we were predominantly getting full MARC records of good quality, either direct from suppliers, or from OCLC, and were managing updates to collections quite effectively. However, for most material in other formats (with notable exceptions, such as the Henry Stewart Talks, a collection of streaming video presentations), it was a different story, and in a number of these cases MARC record provision was largely non-existent, resulting in the omission of content from the catalogue.

In lieu of receiving full MARC records for our remaining collections (largely journal collections), we created our own, either manually creating them one by one, or more often using the MARCmaker function available on MARCedit. This is a fairly efficient function to generate MARC files from excel spreadsheets (which were and continue to be provided by suppliers for most collections). The records thus created were, however, significantly more basic than full MARC records. Additionally, journal records in particular require regular updating in a way that records for ebooks generally don't, as journals change their titles, or publisher, are merged or superseded, and are subject to fluctuating availability of content and coverage according to changes in provision as part of packages provided by suppliers. As a result, records needed regular updating, or replacing with fresh loads from updated spreadsheets using the MARCmaker function. This was found to be a less than desirable situation, requiring significant manual intervention to create and maintain collections using less complete MARC records than we were receiving for our ebook collections.

Looking for a solution to this process, we decided to go for the monthly MARC record download service provided by EBSCO. Predominantly planned for journal collections, the procedure involved using a suite of predominately EBSCO products to produce records on a monthly basis. We started using EBSCO's ERM essentials to manage our electronic subscriptions, which then fed into the A-Z database which we switched to (from SFX) to manage link resolution data, which was thus facilitated by EBSCO's Linksource.

The download used the data from the A-Z database to produce MARC files, which we fed into our LMS, and then also into EDS (EBSCO Discovery System) with the nightly upload we use for all our MARC records, where they form the primary record set for searches.

Not just journals

Although the service was primarily marketed as providing records for journals, which was what we primarily wanted to use it for, we received records for the entirety of our holdings as represented on the A-Z database, which is to say the entirety of our e-resource collections. This was not what we were expecting!

The MARC records provided for journals were CONSER records. CONSER was a project managed by the Library of Congress to create full, high quality MARC records for journals, which up to that time were simply not generally available. So these records were high quality records, comparable with the records we had for our books and ebooks. For any material outside of journals, however, the records were not of an equivalent standard, being very basic in nature. These basic records contained less detail than was contained on the A-Z database (so presumably were not created through effectively converting all the detail from that database). They also often contained less detail and were of a lower quality than the records we had previously been creating either manually or using the MARCmaker function.

However, even though these records were of low quality, the method of receiving them (through a relatively low maintenance MARC download) was considerably more efficient than the processes we had been using previously to create records where full records weren't available, and although the records were sometimes lower quality, the extent to which currency could be easily maintained was felt to make up for a relatively slight decline in quality, so in such cases it was decided to use the process not just for journals but also most other electronic material we couldn't get full MARC records for.

For those collections we were receiving full MARC records for (predominately ebook collections) it was decided to retain our previous practice of receiving full records using pre-existing practices. This was not an easy procedure, however: There is no direct interface for customers to switch MARC provision on or off for individual collections or packages. Instead, the procedure for preventing MARC provision of particular collections involves my emailing our EBSCO rep for the service with any packages we don't want to receive records for, who then switches off record provision for those collections. There is also no interface for us to check which records we have set to receive records for or not. Upon realising we were receiving records for our entire collection, I sent a list of all those collections we wished excluded from the procedure. However, as time went on, we realised that not only did we have to inform EBSCO of any new collections we didn't wish to receive records for, but also if we selected extra subject collections, updated to any new year's batches for packages, that we didn't want records for them either. Enquiring as to why this was, it was found that at EBSCO's backend MARC provision was switched off not at the collection or platform level but at the package level, which is often set up individually according to each subject, year, collection or subcollection entry, so even if we informed EBSCO to stop sending records for any collections in a package, new entries created for other subcollections on the same platforms would not be automatically updated with this block. This results in duplicate basic records regularly re-occurring on our catalogue, an issue we still haven't fully resolved.

Evolution

We decided to switch to this new process for managing our collections in August 2012, at which point we immediately stopped our previous approaches to managing journal collections manually and using MARCmaker. This was perhaps a little optimistic, as it turned out – the process was delayed by many months by repeated departures of EBSCO reps, and periods waiting their replacements (2012-14 was a time of tremendous position and employer switching in the resource supply industry, for some reason). Then, when trying to firm up the procedure, we had further problems in implementing an effective load table to import the records to the catalogue.

In the end, we went live in January 2014 using a less than perfect procedure using pre-existing load tables.

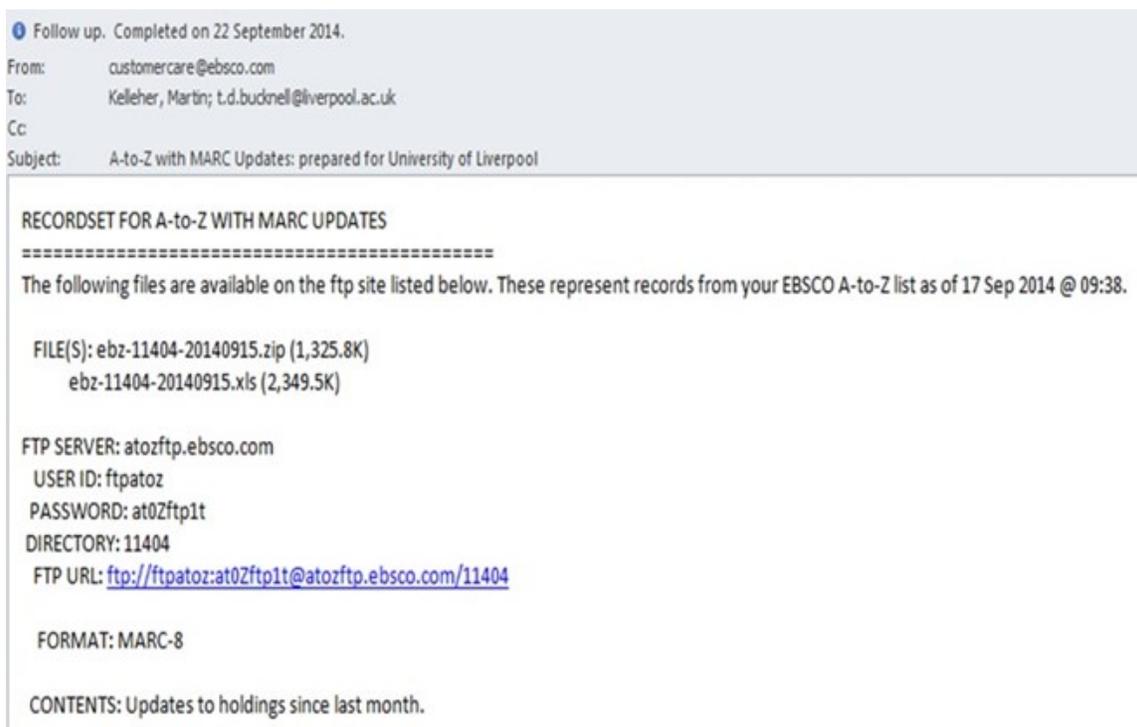
After some experience using the procedure, we found a number of changes were required, not least a more effective load table modified from a pre-existing load table. I co-ordinated the switchover to occur in February 2016. This was undertaken so as to coincide with EBSCO switching from the A-Z database to Holdings Linked Management as the system to manage link resolution (and the MARC download and Publication Finder), and also a switch of the authentication system we were using, from a locally hosted EZproxy to an OCLC hosted EZproxy – Shibboleth hybrid, both processes which required wholesale changes to records.

The new process was a tremendous improvement over the original, but still had some problems and made my own workflow more time consuming. We had a review of the process in 2018 and in May 2018 I implemented version 3 of the load process, with a more fully bespoke load profile. This update was instigated partly because I had received training in load profile maintenance and also in reaction to a problematic and time consuming load resulting from errors in the files sent by EBSCO at the end of 2017.

Blow by blow (with revisions)

At this point I'll give a run through of the process, during which I'll detail changes as they have occurred during the major revisions. The sequence will be numbered throughout the rest of the paper to make the continuity more traceable between explanations.

1. I receive an email from EBSCO:



2. I then download the file from our folder on EBSCO's ftp site:



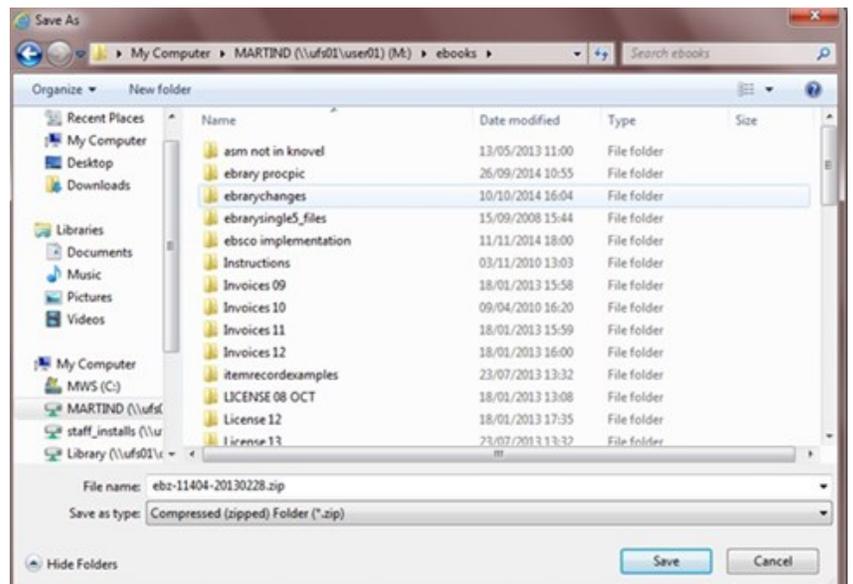
FTP directory /11404 at atozftp.ebsco.com

To view this FTP site in Windows Explorer: press Alt, click View, and then click **Open FTP Site in Windows**

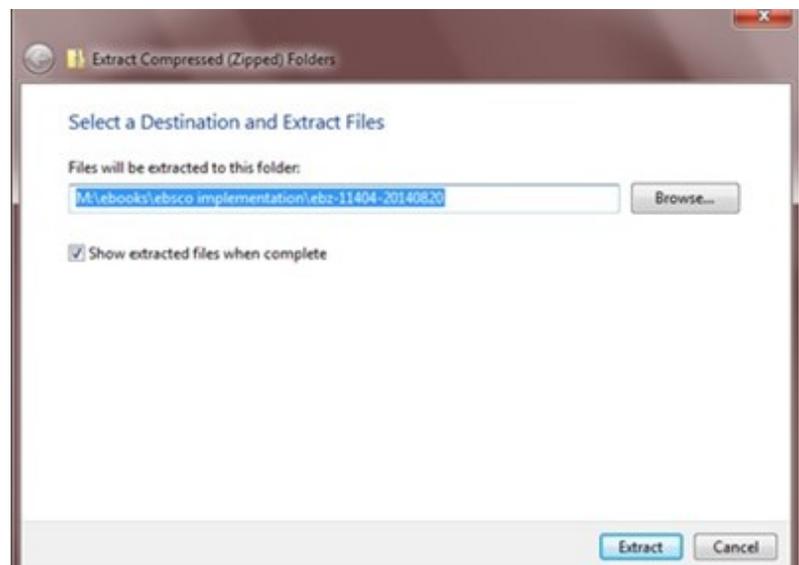
[Up to higher level directory](#)

Date	Time	Directory	File Name	Size
10/20/2014	08:23AM	Directory	.	
10/20/2014	08:23AM	Directory	..	
01/25/2013	12:00AM		ebz-11404-20130123.zip	2,044
01/28/2013	12:00AM		ebz-11404-20130125.zip	5,387
01/28/2013	12:00AM		ebz-11404-20130128.zip	96,579
02/28/2013	12:00AM		ebz-11404-20130228.zip	4,931
03/06/2013	12:00AM		ebz-11404-20130306.zip	4,998
09/17/2014	11:08AM		ebz-11404-20130325.csv	10,198,776
09/17/2014	11:08AM		ebz-11404-20130325.xls	10,385,076
04/19/2013	12:00AM		ebz-11404-20130419.csv	126,081,924
04/19/2013	12:00AM		ebz-11404-20130419.zip	71,794,629
05/17/2013	12:00AM		ebz-11404-20130516.csv	20,234,312
05/17/2013	12:00AM		ebz-11404-20130516.zip	10,085,310
07/17/2013	12:00AM		ebz-11404-20130717.xls	3,981,824
07/17/2013	12:00AM		ebz-11404-20130717.zip	3,564,352
08/18/2013	12:00AM		ebz-11404-20130816.xls	20,634,624
08/18/2013	12:00AM		ebz-11404-20130816.zip	61,238,236
09/04/2013	12:00AM		ebz-11404-20130904.xls	2,015,232
09/04/2013	12:00AM		ebz-11404-20130904.zip	24,517,937
09/17/2014	11:14AM	Directory	ebz-11404-20130916	

3. I save the file to file to my PC:



4. And I unzip the file:



Where the action is: From more is more to less is more

The files as provided by EBSCO initially included (according to our original requirements)

- A file of records for new resources,
- A file of records for updated records for pre-existing resources including improvements and changes to access and..
- A file for records to be deleted because they've been removed from those collections.

We were provided with 2 of each of these, 1 each for serial and monograph records, initially planned to assist with scoping subsequent to loading. This division of files and what I then do with them has changed through the different versions of the load process.

Name	Date modified	Type	Size
Load schedule 1: ebz-11404-20170421.xls	15/05/2017 11:37	Microsoft Excel 97-2...	
Load schedule 1: ebz-11404-20170421-Add.mrc	15/05/2017 11:37	MRC File	
Load schedule 1: ebz-11404-20170421-Addm.mrc	15/05/2017 11:37	MRC File	
Load schedule 1: ebz-11404-20170421-Chg.mrc	15/05/2017 11:37	MRC File	
Load schedule 1: ebz-11404-20170421-Chgm.mrc	15/05/2017 11:37	MRC File	
Load schedule 1: ebz-11404-20170421-Del.mrc	15/05/2017 11:37	MRC File	
Load schedule 1: ebz-11404-20170421-Delm.mrc	15/05/2017 11:37	MRC File	
Load schedule 2: ebz-11404-20170421.xls	15/05/2017 11:37	Microsoft Excel 97-2...	
Load schedule 2: ebz-11404-20170421-Add.mrc	15/05/2017 11:37	MRC File	
Load schedule 2: ebz-11404-20170421-Addm.mrc	15/05/2017 11:37	MRC File	
Load schedule 2: ebz-11404-20170421-Chg.mrc	15/05/2017 11:37	MRC File	
Load schedule 2: ebz-11404-20170421-Chgm.mrc	15/05/2017 11:37	MRC File	
Load schedule 2: ebz-11404-20170421-Del.mrc	15/05/2017 11:37	MRC File	
Load schedule 2: ebz-11404-20170421-Delm.mrc	15/05/2017 11:37	MRC File	

Name	Date modified	Type	Size
Load schedule 3: ebz-11404-20180720.csv	20/07/2018 11:51	Microsoft Excel C...	21,705 KB
Load schedule 3: ebz-11404-20180720-Add.mrc	20/07/2018 11:51	MRC File	7,859 KB
Load schedule 3: ebz-11404-20180720-Chg.mrc	20/07/2018 11:52	MRC File	76,547 KB
Load schedule 3: ebz-11404-20180720-Del.mrc	20/07/2018 11:52	MRC File	844 KB

Load schedule 1

The initial load schedule resulted from our not having an effective load table for updating pre-existing records with updated versions. The result of this was that in cases where records required updating we had to delete them and replace them with new records rather than simply overwriting them. As a result, we had to overwrite such records along with other files to be deleted, then reload them afresh with the new records. The problem with this load was that there was no way to retain information between old and new records, and, since there was a need to modify some of EBSCO's records to delineate according to more standard library format terminology (primarily for scoping on searches), this manual process had to be repeated for each fresh updated version of each record, which considerably increased the manual aspect of record management. Also, since sometimes links used by users, staff, and recorded on the reading list we use (TALIS Aspire) tie in with particular system record numbers (Sierra Bibliographic numbers), these links would break on a regular basis, creating more work. Finally, the process generated a considerable number of deleted records on the catalogue which would clog up the database while waiting to be deleted from not only our local systems but also the various union catalogues we exported to (Worldcat, COPAC and SUNCAT), so a reduction in their proliferation was highly desirable.

Load schedule 2

I figured out a way to use a pre-existing load table created for loading batches of library records. A slightly amended version of this table facilitated version 2 of the load schedule. This process resulted in a much more orthodox three-way load, loading delete files in one load, change files separately to overwrite records requiring updated, and only new files afresh. By doing this we were loading the files in the way they were created for. This load table meant that because some fields were retained it was possible to keep a record of any manual changes that we'd implemented, seriously reducing manual workload to rationalise EBSCO format classification. It also highly reduced the turnover of records, and removed the issue of broken system number based links. The load table was still not perfect however: Some fields would be retained that were to be replaced by incoming fields (the 830 and 856 fields), which could be managed because the fields required some global updating to become fit for purpose, so I could identify and delete pre-existing entries to be replaced and then globally update incoming fields to their correct formats. There was a critical risk that if I didn't undertake this process in the correct order, however, I'd end up with multiple often duplicating entries between which it was impossible to tell the latest version, resulting in the need to redo the complete load. This critical risk, together with other improvements to record quality, meant that the time taken may have reduced various manual elements, but increased the time taken and sensitivity of my own processes.

Load schedule 3

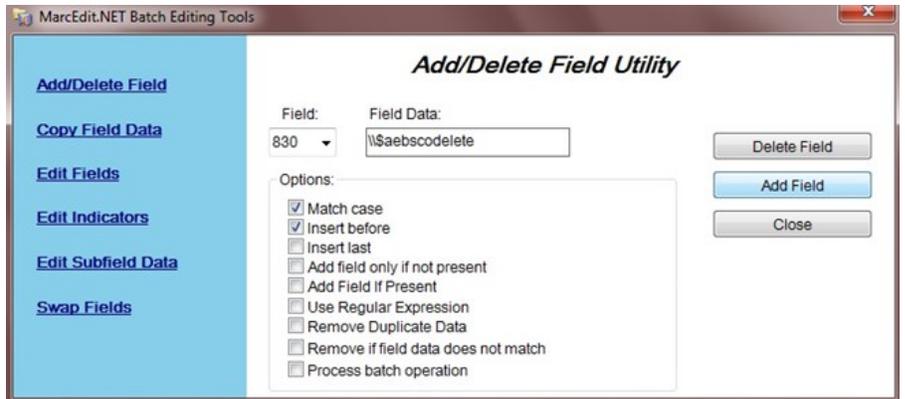
Subsequent to the implementation of load schedule 2, I underwent training in load profile maintenance myself. A few months after this (December 2017), the monthly load included significant errors which took many days to rectify, which acted as an impetus to undertake a considerably more efficient load. Given the remit to fully review the process, I came up with a number of suggestions, one of which was to stop splitting the files between monographs and serials, which I reconfirmed the EBSCO files did not do in a useful way for subsequent scoping on the LMS. This reduced the number of files received from 6 to 3. Since the problem with the December load involved new and changed files being muddled, I decided to conflate those 2 loads into a single file, and simply perform an overwrite load which would create new records where no matches were found. This new load also involved the use of a bespoke table which only retained information in the 084 field (which I used to keep format classification details in) and the 591 field (which retained any local notes we wanted to add) in overwritten records, removing the critical risk element present in version 2 of the load process. This meant I was only using 2 files in the load down from 6, and could overwrite the same records multiple times without worrying about having to perform global updates between loads.

Blow by blow part 2: MARCedit and data exchange

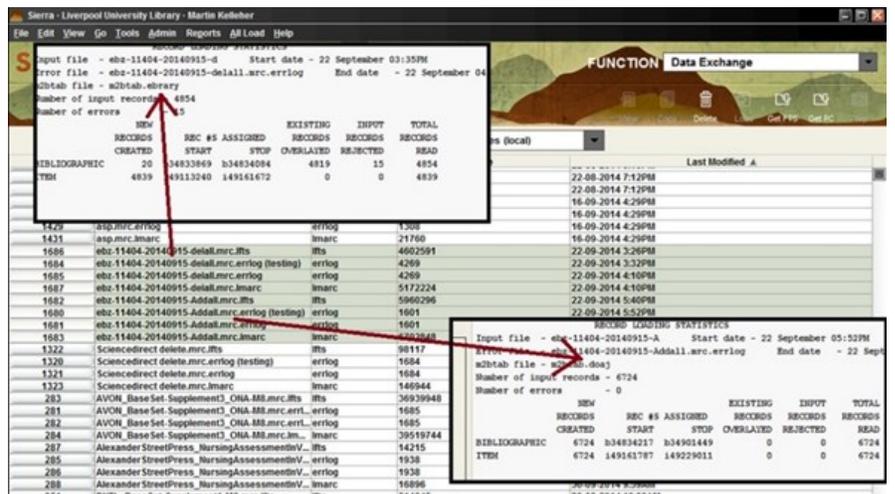
5. The next stage, then, is pre-processing. In the latest version of the load I use the MARCjoin function on MARCedit to combine the add and change files and perform global updates on the resulting file that can't be undertaken subsequent to loading on the LMS.



6. I also add in a series entry to use as a handy reference on the records in the deletion file:

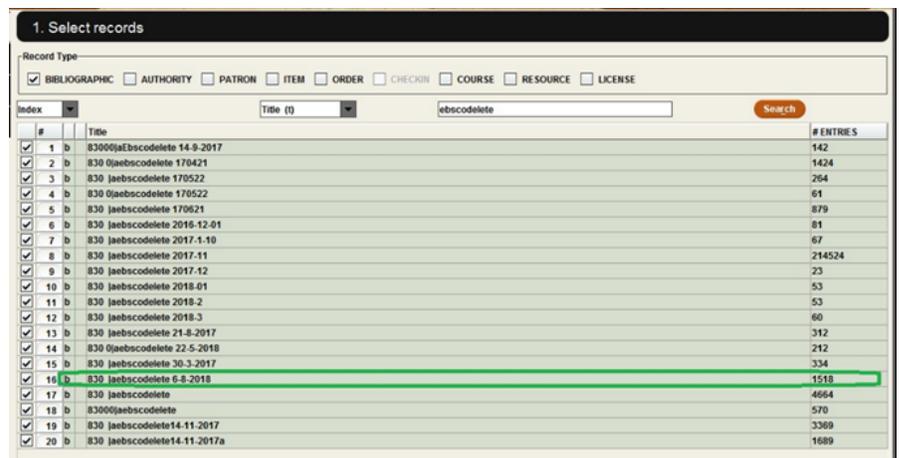


7. Then I load the files onto the LMS (Innovatives Sierra) using Data exchange



8. Deletions

I then run the deletions list first, which I pull up using the inserted tag series entry. A rather basic change to the procedure implemented as version 3 of the load was to Date these deletion fields. Otherwise, if there are deletions still lingering on the catalogue from previous loads, they would be summoned up by the tag field as well, which, if they're remaining on the system for significant periods of time, can seriously slow down processing the deletion process.



9. I next load the new and changed titles.

The full records – version 1

The full CONSER records we receive for journals are as high quality as we were hoping for, containing publication history, alternate titles, title relationship linking, subject headings and all kinds of information we had not previously detailed on our journal records.

However, there were some issues we experienced which required changes in subsequent revision.

Title Journal of accounting, auditing & finance [electronic journal].

Published 2011: Thousand Oaks, CA : Sage Publications
1977-<1980> [Boston, MA : Warren, Gorham & L
<1980>-2010 [Santa Barbara, CA : ABC

Later Publisher

Link to:

00 View online via Business Source Complete (1986 to present)

60 View online via Sage Premier 2014 (1999 to present)

Copies

LOCATION	CLASS NO	STATUS
Electronic Resource	ELECTRONIC JOURNAL	NETWORKED

Frequency Quarterly

Series Online access: EBSCO Business Source Complete
Online access: Sage Sage Premier 2014

More Information Sponsored by the Vincent C. Ross Institute of Accounting Research of New York University, <fall 1985->
Sponsored by Leonard N. Stern School of Business

Subject Finance -- Periodicals.
Accounting -- Periodicals.
Auditing -- Periodicals.

Other Author Vincent C. Ross Institute of Accounting Research.
Leonard N. Stern School of Business.

Other Title Journal of accounting, auditing and finance
JAAF

ISSN 2160-4061 0148-558X

Link Journal of accounting, auditing & finance (Online) 2160-4061 (OCoLC)60620562
2011201672

A Table of Contents is unavailable for this title:

a) ISSNs were combined on a single field, using the \$y subfield. This caused matching problems for enhanced content we'd set up to be provided by Syndetics.

b) We'd initially decided to set up the links as EBSCO Linksource links. This was decided in order to facilitate dynamic changes to links between loads without updating the record, since we could rely on EBSCO to update the link on their database, which the existing linksource link would then direct to. However, this advantage was found to be outweighed by several disadvantages. Firstly, we only realised after implementing the procedure that these links were specific to individual package activations on EBSCO. This meant that if we switched from 1 package to another, often for the same package as managed for 1 years subscription to another (e.g. from the 2014 version to the 2015 version of a subscription deal), the previous deactivated link would cease to function, resulting in dead links on the catalogue until the next MARC update unless we left multiple entries for multiple years active on the A-Z database which obviously led to extra administration. Additionally, since Linksource links were also used on our discovery system, this would lead to a single point of failure situation if Linksource ever went down, as occasionally happens.

c) Dates in the publication field. Our previous records for journals hadn't previously utilised the standard practice of including dates in the 260/264 publication field. We found that when this element started appearing on our systems as a result of being on CONSER records our users found them misleading. The date in the publication field for most materials usually describes the date of publication of holdings, so the journal practice of detailing the full publication range of the journal as published *rather* than holdings was found to mislead our users, even though the correct dates of accessible content were displayed in links.

The full records – version 2

So, when it came to implementing a revised version of the load I amended our practices. We began splitting the ISSNs into separate 022 fields during the MARCedit pre-load process, switched the \$c subfield in the publication field to an invalid field (\$7), which not only removes it from display but also from ranking (which was also negatively impacted by the dated entry), and finally we requested the 856 fields to be sent including direct links rather than Linksource links, all changes of which were co-ordinated as part of the switch over to version 2 of the load process.

Annotations in the screenshot:

- Red arrow: Separate ISSNs to facilitate Synthetics Enhanced content
- Orange arrow: Tagging fields for future removal during catalogue export back to EDS
- Blue arrow: Obfuscation of the publication date
- Pink arrow: Direct links

Not quite full records

Record details:

- Title:** The 19. of Iuly. In this weekly newes is related the great difference or vnlikelihood of agreement between the Emperour and Bethlem Gabor. : With the coming downe of the Marquis of Iaggerendorf, wit [electronic book].
- View this e-book online via Early English Books Online (EEBO) (1622)**
- Copies:** 60
- LOCATION:** Online
- CLASS NO:**
- STATUS:** AVAILABLE ONLINE
- Series:** Online access: Proquest Info & Learning Co Early English Books Online (EEBO)

Having given a run through of the full records, it's worth contrasting the quality and extent of the CONSER records with that of the basic records we were otherwise receiving for non-journal records. Above is a fairly typical example of a basic record - No information beyond the title, the link, and a series entry detailing platform and package. Some basic records are provided with more details than this, and the extent of that detail has slightly increased over the years, but essentially they remain very basic records. However, in some cases they remain the only records we can get for some material, either because MARC records simply aren't supplied for many packages, or, as in this case (Proquests Early English Books online), the price of the records has always been considered prohibitively high, so we've never bought them, leaving these basic records the only records available.

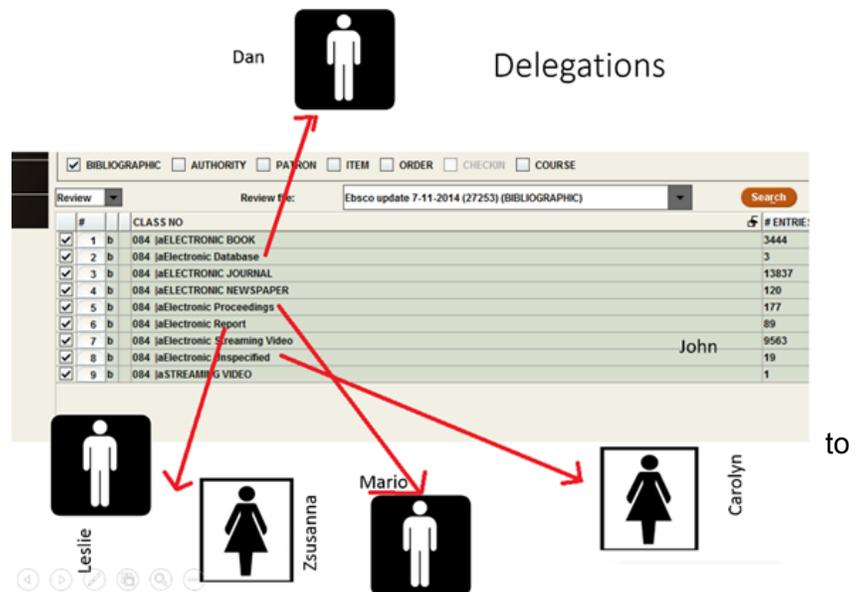
645...b51478687	Old	y - "084	ja201807ELECTRONIC Journal"
	New	c - "084	ja201807ELECTRONIC Journal"
645...b50229011	Old	y - "084	ja201807ELECTRONIC Journal"
	New	c - "084	ja201807ELECTRONIC Journal"
645...b31104484	Old	y - "084	ja201807ELECTRONIC Journal"
	New	c - "084	ja201807ELECTRONIC Journal"
645...b35859477	Old	y - "084	ja201807ELECTRONIC Journal"
	New	c - "084	ja201807ELECTRONIC Journal"
645...b50229035	Old	y - "084	ja201807ELECTRONIC Journal"
	New	c - "084	ja201807ELECTRONIC Journal"
645...b39760649	Old	y - "084	ja201807ELECTRONIC Journal"
	New	c - "084	ja201807ELECTRONIC Journal"

11. Corrections

After performing the bulk of global updates, I then run through the less common formats, which seem to be most frequently those including errors (especially material attributed as being streaming audio and video, for some reason), correcting and reassigning their format. I then compile and send these corrections in an email to EBSCO, so that they can update their systems to ensure the same errors aren't perpetuated, especially in terms of being reapplied to fresh updates of records. In this same email I request the removal of records from packages I notice we're getting records for that duplicate superior records we source elsewhere.

12. Delegations

Finally, after performing the majority of the work myself, I pass on any records to be manually sorted by format to library assistants in the subscriptions and metadata sections. These include a) the aforementioned Reports and Proceedings categories, in addition to b) those items whose category of "unspecified" is exactly what you'd expect, so requiring local categorisation, and finally c) databases, which are checked to deduplicate with full records. This latter task remains necessary because EBSCO's processes do not allow us switch off record provision for general database records without also switching off provision of records for content of such databases which we rely on the process for records of, so this remains something we must do locally. The Metadata library assistant also runs through any error reports detailing cases of records which don't successfully overwrite during the load process.



What remains?

What remains with regard to the future of the process remains almost exactly the same as remained the year before, when I gave the version of this presentation pertaining to version 2 of the process.

1. The rather awkward email correction approach required to repeatedly request the provision of duplicating collections remains a problem, perhaps solvable by improving our local processes, or ideally a more accessible version of the process to be provided by EBSCO.

2. Format disparity. Although manual intervention to make EBSCO categories library friendly have become more efficient over the evolution of the process, it remains more time consuming than is desirable. I've considered more efficient ways to delineate during the manual process, but a truly efficient approach to processing these records remains an issue.

3. Finally, for all the advantages the process provides, there are various areas where improvements could obviously be made, and the possibility to replace the whole process with an altogether improved process is ultimately something I'm hoping for.

EBSCO HLM > III Knowledge Base > Skyriver > LMS = Process perfected?

For some time there has been a collaborative project between Innovative and EBSCO to develop a Knowledgebase which can facilitate a more library adapted version of the process. I believe that the processes I've been describing an efficient and effective one, but I'm hoping for some improvements to be made in this succeeding suite of collaborating products. I hope these to include:

1. A streamlined process, especially considering we already utilise both products as part of our services already.
2. Skyriver to provide significantly superior records for non-journals, given it's more standard library metadata basis. I'd also hope that the Skyriver database would provide more complete datasets than EBSCO, which can be particularly incomplete for ebook collections.
3. At least the same extent of coverage, in which case I would consider consolidating the provision of records to include some or all that we currently source from OCLC or direct from the providers.
4. Finally, resolution to a problem I haven't so far mentioned. EBSCO records provide multiple links on records for each package a resource is available through, including multiple entries for the same platform if access is available via multiple subscriptions to the same publication. I would like to see links for the same platform to be automatically consolidated, presenting only 1 link per platform with a combined statement indicating the range of coverage across multiple package access rights if relevant.

And so lie my hopes for the future evolution of the process by our providers. However, I wouldn't rule out the option of going elsewhere: I'm a member of one of the groups feeding into the development of the National Bibliographic Knowledgebase, which may provide some kind of service to replace that provided through our current approach: There is also Folio, an open source product EBSCO has been involved in developing, and there is also the hope that we could use Worldshare Services to manage our records further. There is also the possibility of a move to another system altogether, perhaps Alma/Primo as a replacement to our current suite of systems. Or the future may hold other possibilities – only time will tell.

Biography

Martin began his career in short term, mostly project managing, positions. He moved into desk services, cataloguing and then a dual role between cataloguing and electronic resource management. His current position is as Metadata Manager at the University of Liverpool

martind@liverpool.ac.uk

Zines, while still under-collected, have become increasingly visible in libraries and archives. In libraries in particular, zines pose special cataloguing challenges. Those challenges illuminate broader implications for librarianship as they forefront the value of cataloguing and metadata in the field. The challenges of zines, as will be discussed, demonstrate why high-level expertise and judgment are required of metadata practitioners as well as, crucially, a willingness and ability to incorporate the interests of multiple stakeholders in libraries' collections, including creators and subjects of library material.

Beginning with the consideration of library metadata in general and its ramifications for future generations, if we do not prepare for the structured discovery and use of cultural heritage data in the future, there is a sense in which there will be no future. At least, there will be no future representation or access to the world's cultural heritage, which is probably not a future worth having.

The case that will be made here – for the value and necessity of metadata work as a profession – will be made in terms of ethical parameters: that the future, and the quality of the future for our fellow human beings, depends in part on us and the work that we do. Namely in the realms of rights to privacy and the enabling of social change: greater equality, equity and diversity in representation.

If a good future hinges on good metadata, then what is good metadata?

Good metadata adheres to content and encoding standards. It is skilfully created and harmonized. It is shareable and interoperable. However, a larger question remains: in the context of professionally practicing in metadata, what is good? How do we do good with our metadata?

As librarians are beginning to question and reject conceptions of libraries-as-neutral (“Are Libraries Neutral?,” 2018), new perspectives emerge on the value and import of metadata and on how it can have good and/or sinister consequences. Determining how to wield metadata for good may provide fresh rationale for how much metadata deserves to be supported and resourced.

So, what is good?

Surely, if perhaps simplistically, it entails identifying and doing the “right thing,” being humane, reducing harm. These kinds of considerations have not arisen often enough in metadata circles. Questions like: When we consider who we are serving in our metadata work, who is being left out? And what are the consequences of that exclusion?

Zines are a great example to illustrate answers for the question of how do good with metadata. With growing interest in zines in cultural heritage institutions, there is growing awareness of the challenges that zines pose, particularly in metadata.

Why do these challenges exist? In large part because zines are not made for us. They often have a limited range of audiences. Take the example of zinester Michelle, quoted in Fox & Swickard (forthcoming), who explicitly seeks to exclude family and employers as potential audiences.

“My main concern is that my family will somehow find out my sexual orientation, which I intend to keep private. I also worry about whether or not my critiques of my workplace will somehow end up in the hands of my boss. I also don’t like many people knowing about my past drug-use, for fear of judgement, so I like to keep that information private as well.”

Zines often are not made for libraries either, nor bookstores. They are made for other information networks outside of and often intentionally apart from everyday cycles of information exchange and consumption. Zines are made for the zine community, which is a varied group of communities, as communiqués, symbols, calls to action, the fabric of networks of support and care for marginalized people, or as anonymous acts of self-care for someone who simply needs to express something.

All of this creates some incompatibility with how we approach metadata creation, in which we generally assume that the creator should be identified with their work and that the work should be discoverable. Based on what librarians working with zines have come to know, neither of these may always be the case for zines (Berthoud et al., 2015).

And yet zines are also material records of these marginalized communities. They are both primary sources that should co-exist with and challenge other library resources and exciting means for the communities represented in zines to see themselves rightfully represented in cultural heritage collections. It is really important that zines are a part of posterity.

Balancing these opposing considerations yields a consensus something like this: having zines in libraries is an imperfect fit, but it is important enough to make it work. The incompatibilities remain. Dealing with those incompatibilities helps illustrate how we ought to practice good metadata.

The sometimes-opaque nature of zines sets us up to be metadata detectives. They might lack basic identifying bibliographic information like author, title or date. We cataloguers take pride in our ability to track this information down. Fox and Swickard (forthcoming) critique this tendency to sleuth.

“Catalogers need to understand the very real danger posed by sharing [information] about, for example, street artists who are protesting police violence or city inaction against economic hardships, or transgender artists who do not want previous art created under a dead name connected to their current art.”

Just like the example of Michelle, zines’ intimations can create unexpected risks for zinesters when they end up in libraries. Cataloguers’ diligence to overcome obfuscation by sleuthing out data elements can exacerbate those risks, exposing sensitive information to wider readership or, after cataloguing, to dissemination via online networks.

Consider zine metadata on networks like WorldCat and local library bibliographic data being transformed into linked data. That metadata could be disseminated and stored or cached across the internet. Such large-scale information networks are where the tensions of doing metadata work with zines will play out, with both costs and benefits.

Consider the upsides. Networked zine metadata has revolutionary potential as seeds of dissonance and change. It represents a potential re-coding of dominant knowledge systems, a remapping of epistemologies by infiltrating the semantic web with new, more inclusive, transformational ways of knowing. Kate Eichorn (2013, pp. 152–153) connects zine cataloguing to Donna Haraway’s *“Feminist cyborg stories”* – *“recoding communication and intelligence to subvert command and control,”* making *“previously unimaginable identities ... that resist the prevailing binary code ... visible.”* Eichorn sees zine cataloguing as *“part of a larger epistemological project”* in which re-inscription of zinesters’ ways of knowing into machine readability *“may hold even greater potential for social change than the act of media transfer itself”* (i.e. digitization).

However, these seeds of change are also the seeds of exposure. All of this potential-laden visibility is accompanied by an increased vulnerability insofar as linked data remains uncertain in its capacity to be changed. Linked data can certainly be changed at its source, but once disseminated across the web and cached by third parties, it is not at all clear that it can be universally changed or recalled.

So what do we do? There are nascent conversations about ethics in cataloguing work that deal with issues like these, in part by advocating for building such considerations into our professional practice. Take for example, the *Zine Librarians Code of Ethics*, prepared by participants of the North American-based Zine Librarians Interest Group.

“To echo our preamble, zines are ‘often weird, ephemeral, magical, dangerous, and emotional.’ Dangerous to whom, one might ask? ... prospectively ... zinesters themselves. Librarians and archivists should consider that making zines discoverable on the Web or in local catalogs and databases could have impacts on creators – anything from mild embarrassment to the divulging of dangerous personal information. [Therefore] [z]ine librarians/archivists should strive to make zines as discoverable as possible while also respecting the safety and privacy of their creators.” (Berthoud et al., 2015, p. 16)

Caswell & Cifor’s (2016) concept of radical empathy in the archives profession has been an inspiration for subsequent conversations in the zine librarian community, such as those at the Zine Librarians UnConference 2018 (“ZLuC 2018 MSP Notes,” 2018). They argue for a shift away from a rights-based approach to archival work and a detached professionalism, towards one founded on a feminist ethics of care that acknowledges webs of mutual affective responsibility between archivists, record creators, record users, the subjects of records and the community at large. This approach makes marginalized communities “*not just a target group of users, but central focal points in all aspects of the archival endeavor.*” (Caswell & Cifor, 2016, p. 24)



Figure 1: image courtesy of the author, pp.3-4 of: Wooten, Kelly (2018). Radical Empathy in the Archives. <https://tinyurl.com/wootenRadEmpathy2018>. Licensed under Creative-Commons BY-NC.

This graphic is taken from a zine created by Kelly Wooten of the Sallie Bingham Center for Women's History and Culture at Duke University's Special Collections Library. It illustrates the four affective responsibilities laid out by Caswell & Cifor (Wooten, 2018, pp. 3–4). Caswell & Cifor also figure directly into Fox & Swickard's main argument for the responsibility of library cataloguers to recognize the needs of the subjects of authority records in traditional library catalogues; that cataloguers should "*take the time to recognize and empathize with the persons [being described]*" in authority records (Fox & Swickard, forthcoming).

These evolving conceptions of how to approach zine librarianship and cataloguing involve responsibilities along a continuum of institutional roles, from zine collection curators to cataloguers. They entail a broad range of questions about how professional practice could evolve – not just with zines, but with any information resources that implicate marginalized or potentially vulnerable populations. These include questions such as:

1. What if it became common practice to seek consent from the subjects or creators of zines before we include them in our collections?
2. What if we included creators and subjects of zines in our description and organization work – ensuring that we include or exclude the information that will adequately represent and protect them?
3. And how might we develop the standards and best practices that guide our work to acknowledge and help us navigate these mutual affective responsibilities?

However, there is meaningful conflict between user expectations and zinester (i.e. creator) needs. Though this paper and most zine librarians advocate for the often under-considered creator's point of view, we must think carefully about what we are giving up when we make decisions to benefit or protect the creator. Cataloguing and metadata have held user needs at their centre from Cutter to Ranganathan to FRBR. Extending our responsibilities to include creator needs reveals the fundamental tension between the interests of these two groups. Acknowledging that tension and beginning to grapple with it is a necessary burden in the evolution of the profession, one that will require great skill and judgment.

Put another way, under a new, broader sense of affective responsibility, work with zines or other like materials becomes weightier and more complex. That weight and complexity calls for a more nuanced approach, new requirements for professionalism, an ability to wrestle with sometimes-conflicting ethical considerations. We need high-level expertise in metadata and its lifecycles – understanding where our metadata might end up, and the potential and risks posed. We need contextual knowledge of the zine community or other at-risk communities, a commitment to genuine relationships with them, to be a partner to them, to draw on their support and consultation as we literally encode their lives and knowledge into the cultural record. And, crucially, we need the institutional and professional support and empowerment to operate at these highly skilled, highly complex levels. To do no harm, we must do metadata right, including the concept of "doing the right thing." Metadata operations should therefore be supported and resourced to enable that level of stewardship and responsibility.

Biography

Joshua Barton is Head of Cataloging & Metadata Services, Assistant Head of Technical Services, and Zine Librarian at Michigan State University Libraries. His background is in serial cataloguing, in which he remains active as MSUL's CONSER representative.
bartonjp@lib.msu.edu

References

- Are Libraries Neutral? (2018, June 1). Retrieved November 30, 2018, from <https://americanlibrariesmagazine.org/2018/06/01/are-libraries-neutral/>
- Berthoud, H., Barton, J., Brett, J., Darms, L., Fox, V., Freedman, J., ... Wooten, K. (2015). *Zine Librarians Code of Ethics Zine*. Durham, NC: Zine Librarians unConference.
- Caswell, M. L., & Cifor, M. (2016). From Human Rights to Feminist Ethics: Radical Empathy in Archives. *Archivaria*, 81(Spring 2016), 23–43.
- Eichhorn, K. (2013). *The archival turn in feminism: outrage in order*. Philadelphia: Temple University Press.
- Fox, V., & Swickard, K. (forthcoming). "My Zine Is My Private Life": Reframing Authority Control from Detective Work to an Ethics of Care. In J. Sandberg (Ed.), *Ethical Questions in Name Authority Control*. Sacramento, CA: Litwin Books.
- Wooten, K. (2018). *Radical Empathy In Archival Practice* ([2nd edition]). New Haven, CT: Kelly Wooten. Retrieved from <https://tinyurl.com/wootenRadEmpathy2018>
- ZLuC 2018 MSP Notes. (2018, July 14). Retrieved November 30, 2018, from <http://zinelibraries.info/wiki/zluc-2018-msp/notes/>

Introduction

This paper builds upon an earlier one written in 2015.¹

Serials are part of a larger world of continuing resources, defined by the ISSN Manual as “*a publication, in any medium, that is issued over time with no predetermined conclusion and made available to the public*”. This includes e.g. certain types of websites and blogs. In RDA, the term continuing resources is not used, only referring to serials. Therefore, in this paper I will use the label serials, although it will be impossible not to refer to continuing resources at all!

Serials can be complicated, to quote Ed Jones from his 2013 book *RDA and serials cataloguing* : “*a serials catalog record is never finished, not even when you’re absolutely sure it’s dead, not even when you’ve ritualistically driven a stake through its heart, pounding with all your might*”.² It is this fluidity of structure, propensity for change and evolution, which makes serials interesting and challenging. Fitting them into the structure of any standard that does not accommodate their volatility is difficult.

This paper gives an overview of the main standards how they have developed and looks at some potential outcomes of current developments.

The ISSN Network and standard

The ISSN standard has its origins in work between IFLA and UNESCO in the late 1940s and 1950s. By the 1970s, UNESCO had two programs relating to libraries and information policies. One, UNISIST intended to foster the development of a world scientific information system. The International Serials Data System (ISDS) was initiated under the auspices of this program. One outcome was a specific recommendation to create a global register of scientific journals, advocating a “*universally accepted code applicable to scientific journal titles*”.³ A resolution adopted at the Sixteenth session of the General Conference of UNESCO authorized the Director General of UNESCO “*to foster international cooperation in scientific and technical information*”

An outcome of this was an agreement between the Bibliothèque National de France (BnF) and UNESCO resulting in 1971, in a report by the French librarian Marie Rosenbaum, highlighting the differing needs of the constituencies, ISDS would need to reconcile and introducing the idea of a collaborative two-tier structure for the ISDS of an international centre linked to national centres. It promoted the production of a database including serials irrespective of subject. In 1972, the report was presented at the 38th session of IFLA General Council. ISSN was defined by ISO standard 3297 first drafted in 1972 and adopted in 1975.

1. Bequet, Gaele, Howlett, Louise, Willer, Mirna “Interaction between IFLA standards and other library standards: ISBD, RDA, UNIMARC and ISSN: a long-lasting relationship” Paper presented at: IFLA WLIC 2015 - Cape Town, South Africa in Session 166 - Standards Committee.

2. Jones, Ed 1951- “RDA and serials cataloguing” Chap. 1, p.4 ISBN 978-0-8389-1139-6

3. UNESCO. Abrégé de l’étude sur la réalisation d’un système mondial d’information scientifique effectuée par l’Organisation des Nations unies pour l’éducation, la science et la culture et le Conseil international des unions scientifiques. [online]. 1971. <http://unesdoc.unesco.org/images/0013/001356/135601fo.pdf> [5/01/2015]. p. 46-47.

A report commissioned by the ISDS International Centre in 1992 concluded that, the information community at large did not make any clear connection between ISDS and the ISSN standard, consequently ISSN needed including in the name of the International Centre (this change occurred in 1994).

The ISSN standard has had several revisions and is undergoing another (2016-).

International Standard Bibliographic Description (ISBD)

ISBD designed by IFLA's cataloguing section to serve "as a *principal standard to promote universal bibliographic control...*". ISBD(S) – serials, initially published in 1974 when the only available frame of reference was ISBD (M) - monographs. Achieving compatibility with this and harmonizing with ISSN's serial specific guidance proved problematic. The issues raised led in part to the general framework ISBD (G) published in 1977, followed by the first standard edition of ISBD(S) this conformed very closely to ISBD (G)'s structure. Albert Mullis⁴ commented that by 1981 "although perfect and absolute compatibility has not been achieved there is now a workable harmonization."⁵ This "workable harmonisation" continued through the early 2000s. ISBD(S) underwent a significant revision in 2002 and was renamed as ISBD (CR) – continuing resources; a consolidated edition replacing all the individual publications was published in 2007. This has had several revisions including the current one, which aims to align the ISBD element set more closely with that of LRM.

Anglo-American Cataloguing Rules (AACR) and the move to Resource, Description and Access (RDA)

AACR developed out of the long established but separate historic cataloguing codes in use in the UK and the USA. AACR had its first publication in 1967 but with separate North American and British texts. During the early 1970's work was done on aligning with ISBD. The first consolidated edition of AACR2 being published in 1978. Other revisions followed, the last in 2002 included chapter 12 – "Continuing Resources". Discussions over the next major revision of AACR led ultimately in 2010 to the publication of RDA – Resource Description and Access by the RDA Steering Committee (formerly the Joint Steering Committee for the Development of RDA) underpinned by IFLA's conceptual model Functional Requirements for Bibliographic Records (FRBR).

IFLA - Functional Requirements for Bibliographic Records (FRBR)

IFLA's conceptual models have had a considerable impact. The entity relationship model FRBRer, was published in 1998 focusing on descriptive bibliographic data. Generally very well received the original report itself acknowledged it did not represent serials very well stating, "in particular, the notion of "seriality" and the dynamic nature of entities recorded in digital formats merit further analysis."⁶ The situation in 2010 when RDA was published was broadly unchanged; little substantive work had been done on serials and seriality. A cause for concern to those closely involved with serials.

4. Albert Mullis was Head of the Serials Office, Bibliographic Services Division, The British Library and served on the Governing Board of the International ISSN centre, 1981-1988 as well as editing the ISDS manual.

5. "ISDS, ISBD(S) and AACR2: Divergence and convergence" Albert a Mullis *Serials Cataloguing*, 1992 Haworth Press.

6. Section 1.3 Areas for further study p.5 "*Functional Requirements for Bibliographic Records – Final Report*" IFLA Study Group on the Functional Requirements for Bibliographic Records September 1997 published 1998.

The ISSN Review Group the road to PRESSoo and beyond

The ISSN Review Group⁷ was established to respond to the challenging environment experienced by both the bibliographic and the publishing communities, e.g. the implementation of RDA, transition from MARC, harmonization, interoperability and mapping between standards, the FRBRization of catalogues, linked data etc.⁸

RDA led the ISSN review group to re-examine the ISSN manual. They considered that they needed active and formal engagement with other standards groups, RDA Steering Committee (RSC) for RDA, IFLA-FRBR RG and the ISBD RG as well as coding schemas like MARC, to avoid the potential risk to harmonisation between ISSN and the other standards. One example is the protocol signed between the ISSN IC and the RSC originally agreed in, 2015, formalizing a relationship begun in 2011 when members of the ISSN RG attended JSC's meeting in Glasgow as observers. This was renewed in January 2016 (<http://www.rda-jsc.org/sites/all/files/RSC-Chair-13.pdf>).

Analysis of the FRBRer model in the context of serials was a major strand of work for members of the ISSN Review Group, their objectives were:

- To improve the understanding of the FRBRer model from an ISSN perspective and its applicability to continuing resources.
- To contact FRBR specialists in order to present concerns and issues identified.
- To establish internal ISSN “guidelines” for application of FRBR to continuing resources (that could be used, for example, in the framework of the FRBRization of catalogues).

They presented their analysis at the ISSN Directors meeting in Sarajevo (2011) reaffirming that the FRBRer model did not work well for continuing resources and within its present structure the model could not be made to work effectively. With FRBR, concepts increasingly embedded within the bibliographic community and some national libraries or bibliographic networks planning to “FRBRize” their catalogues. How could we “FRBRize” or have discussions based on FRBR concepts when the model did not accommodate serials and other continuing resources well?

To address this the Review Group held discussions with several FRBR specialists (Françoise Leresche and Patrick LeBoeuf from the BnF, Philippe Le Pape from ABES⁹ and Gordon Dunsire current RSC chair) during a technical meeting in Paris, April 2012. The meeting provided promising new avenues in particular, Patrick LeBoeuf presentation of FRBRoo¹⁰ an extension of FRBRer demonstrated it might accommodate the features of continuing resources more effectively (CIDOC-CRM¹¹ are co-responsible with the FRBR RG for FRBRoo - FRBR model based on the Object Oriented modelling).

7. The ISSN Review Groups role is to prepare proposals for modification of the ISSN rules and to liaise with the other bibliographic groups or communities on behalf of the ISSN Network. The work of the group is reported, for preliminary discussions or validation, through the ISSN mailing list and during the annual ISSN Directors meeting.

8. Current members of the group are from: the ISSN International Centre, Bosnia and Herzegovina, France, Germany, Morocco, Poland, Slovakia, Sudan the UK and the USA.

9. ABES – Agence bibliographique de l'enseignement supérieur created in 1994 to implement the union catalogue of Frances higher education libraries - Sudoc

10. FRBRoo Introduction (version 1.0), http://www.cidoc-crm.org/frbr_inro.html. See more on: Working Group on FRBR/CRM Dialogue, <http://www.ifla.org/node/928>

11. “The CIDOC CRM is the culmination of over 10 years work by the CIDOC Documentation Standards Working Group and CIDOC CRM SIG which are working groups of CIDOC. Since 9/12/2006 it is official standard ISO 21127:2006.” <http://www.cidoc-crm.org/>

Subsequently the analysis was presented to the FRBR Review Group at the IFLA Conference (Helsinki, August 2012). This confirmed that issues around the applicability of FRBRer to continuing resources, especially seriality, were not adequately dealt with and although the report of the FRBR Working Group on Aggregates, released in September 2011¹² considered serials as common aggregates, it did not address most of the topics of concern to the ISSN Network.

The FRBR Review Group had no work scheduled regarding continuing resources, so the ISSN RG accepted a proposal from the BnF establishing a working group between them to work on the application of FRBRoo to continuing resources.

The ISSN/BnF Working Group met January to March 2013 and analyzed the semantics of each element in the ISSN manual. Wherever classes and properties were not available in FRBRoo or in CIDOC CRM, new ones were declared forming the core of PRESSoo. A mapping from all of the ISSN data elements to PRESSoo was done in parallel.

In 2013 the ISSN/BnF WGs published version 0.1 of the PRESSoo model the following quote sums up the nature of the problem they aimed to resolve:

“continuing resources pose a particular modelling issue, in that their descriptions do not only reflect characteristics of existing products, but also, as long as the described resource still is being published, the expected characteristics of future behaviour. The main difference between cataloguing a monograph and cataloguing a serial could be expressed as follows: when you catalogue a monograph, you make statements about the past; when you catalogue a serial, you both make statements about the past and assumptions about the future.”¹³

Work progressing PRESSoo now moved to IFLA with a formal IFLA PRESSoo review group established. Version 1.3 of PRESSoo was formally approved by IFLA on the 31 March 2017.

IFLA FRBR Review Group in 2013 had been working on a consolidated model for FRBR and in May 2016 published their draft model. Renamed as the IFLA Library Reference Model (IFLA LRM) it received formal approval in August 2017.

IFLA - LRM and alignment between ISSN and RDA

The LRM enables progress on harmonization between ISSN and RDA. It is recognized that interoperability between them is required because of the frequency that data is exchanged between institutions using both RDA and ISSN instructions and element sets. This is particularly true between ISSN National Centres that have adopted RDA and the ISSN IC. Data from National Centres is ingested into the ISSN Register, the international database that stores and gives access to authoritative ISSN data from all contributing centres.

The decision by the ISBD RG at the IFLA general conference in Wroclaw 2017 to produce a revision of ISBD aligned to IFLA-LRM has the potential to further impact the ISSN manual, traditionally aligned with ISBD adding further impetus to continuing the work on harmonization.

12. Cataloguing Section. Final Report of the Working Group on Aggregates. September 12, 2011, <http://www.ifla.org/files/assets/cataloguing/frbragg/AggregatesFinalReport.pdf>

13. PRESSoo [Extension of CIDOC CRM and FRBRoo for the modelling of bibliographic information pertaining to periodicals] version 0.1 Mar 2013 Editor Patrick Le Boeuf (BnF) p.5

PRESSoo had an impact upon the LRM compared to its predecessors it takes a radically different approach to issues related to seriality. The model has a section dedicated to serials (5.8) recognizing they have characteristics making them complex to model. The LRM agrees with PRESSoo that when cataloguing serials you have to deal with what is known in the present and try to anticipate future behaviour.

Outcomes from ISSN RDA alignment

There is a considerable amount of work ahead for both RDA and ISSN as the change of approach delineated in the LRM will have a significant impact, but there are areas where the two have a common view. Section 5.8 of the LRM model makes several points para-phrased from the recent discussion paper¹⁴ below:

- Serials aggregate several levels of content normally at least serial title, issue, article levels.
- As dynamic resources, the description of a serial work “does not limit itself to a description of the past...”
- It is not possible to predict with certainty that any relationship between serials, that is true now, will hold true in the future.
- Collocation based on the “commonality of content” cannot be applied to serials: It follows that “any serial work can be said to have only one expression and only one manifestation”. This is referred to as the “WEM lock.”¹⁵. A major point, as collocation is a significant outcome of both the FRBR model and subsequently the LRM.
- Many data elements should be relocated to the level of the work e.g. only one element related in RDA to “mode of issuance”, the number of units or carriers should remain at the manifestation level the others would become work elements.

The serial work

No serial work is ever complete until the title has completely ceased, but it was not an idea taken up by the conceptual models until the LRM. In the LRM, a “serial work” is a *publication plan*, i.e. the plan to publish a serial (or any type of continuing resource) with certain features or characteristics. We need to distinguish between those resources that do or do not have a predetermined end, and for the latter to create specific instructions for them. The weight of this falls upon RDA as the ISSN manual is already wholly concerned with continuing resources.

Theoretically, any changes in the plan should generate a new serial work. However, not all changes are significant enough to generate a new work. It is crucial to define what type of change requires a new description e.g. a change of frequency on its own should not generate a new description. RSC and ISSN IC will need to work together in order to provide harmonized general guidance and/or specific instructions for cataloguers.

As both ISSN and ISBD use the label continuing resources to delineate the scope of what’s covered in their instructions it might be appropriate for RDA to reintroduce this label.

14. Coauthored by RSC and ISSN IC and approved by the ISSN RG – “Issues on IFLA-LRM alignment for serials and other continuing resources” RSC/Chair/20 16 June 2018 <http://rda-rsc.org/node/559>

15. Coauthored by RSC and ISSN IC and approved by the ISSN RG – “Issues on IFLA-LRM alignment for serials and other continuing resources” RSC/Chair/20 16 June 2018 <http://rda-rsc.org/node/559P2>. Of discussion paper “Issues on IFLA-LRM alignment for serials and other continuing resources”

The RDA/ONIX framework

Another change will be to the vocabulary used. To define what section 5.8 of the LRM describes the RSC and ISSN IC propose to use vocabulary from the RDA/ONIX framework (ROF).¹⁶ Specifically the extension mode, extension termination and extension requirement.

Extension mode defines whether the resource is integrating, successive or it is not applicable. Those resources where the extension termination is “indeterminate” i.e. has no predetermined end are covered in section 5.8 of the LRM. Another ROF value utilized is “extension requirement”, resources described, as “indeterminate” have an extension requirement of “inessential” i.e. if a subscription is cancelled or it ceases publication the content is still viable and useful. Whereas for an integrating resource described as “indeterminate” but with an extension requirement described as “essential” if it ceased or a subscription cancelled, the resource would lose its value to users its content becoming outdated and limited in application.

Identifier for the serial work

The UNISIST study leading to the development of ISSN, advocated a “universally accepted code” for journal titles each ISSN representing a single serial work, different medium version/s of the same work are assigned a different ISSN, so fulfilling its role as a title identifier. In RDA, ISSN has been seen as only identifying different mediums and been identified with the manifestation entity, not the work.

The development of the LRM has changed this, the “WEM lock” principle says each medium version is a different work in its own right and that in, the WEM “stack” there can only be a single direct relationship between the WEM levels of a serial so an identifier for any of the levels identifies all of them. The LRM considers that ISSN should be the identifier for the work not the manifestation. A significant change leading to further discussions between the ISSN IC and RSC RDA.

The LRM states that collocation based upon “commonality of content” is not possible for serials. It also says, “All relationships between serials can be modelled as work-to-work relationships”. Types of relationships include:

- Those where the original serial work has ceased publication: e.g. mergers, split, etc.
- Those where the original serial work has not ceased: e.g. is published in another medium, another language, a new local edition, etc.

The LRM states that it should be possible to define additional entities as a means of clustering related works. One existing example is the linking ISSN (ISSN-L) which groups together serial works with an “other medium version” relationship demonstrating the potential to define other clusters e.g. one relating all former and successor titles. There are ongoing discussions about the creation of such “family ISSNs” within the working group on the revision of the ISSN standard.

ISO 8 - Presentation and identification of periodicals

The standards referred to so far are integral to the creation of bibliographic description. One other currently under revision is ISO 8 – Presentation and identification of periodicals. Concerned with how key information is presented, it is intended to enable editors and publishers help users of their publications to discover, cite and retrieve information effectively. It has been in place since 1977 but not until now had any revisions. In its unrevised state it is, highly print orientated and completely out of step with the current publishing environment. Revision of the standard aims to improve the consistency and quality of the metadata provided and provide an up to date standard containing best practice irrespective of the media used.

16. RDA/ONIX for resource categorization: <http://www.rda-jsc.org/archivedsite/docs/5chair10.pdf>

Conclusion

Over the last 6-7 years, there has been a significant amount of work carried out by the ISSN community with a wide range of colleagues in different standards groups to raise the profile of serials and continuing resources generally and to improve the level and scope of harmonization of serials treatment within the standards framework. The work is by no means complete but it has proved that whilst serials are complex they are not impossible to model. We may never achieve what Albert Mullis described as “*perfect and absolute compatibility*” but we are getting somewhat closer to it.

Biography

Louise Howlett (British Library) is the Serials Record Handling Manager and Director of ISSN UK. She is also the UK representative on the ISSN Governing Board and currently its Chair and also a member of the ISSN Review Group, UKSG's education subcommittee, and several other standards groups including ISO/TC46/WG07, ISO/TC46/05, and IFLA-PRESSoo.

louise.howlett@bl.uk

Serials. A complicated lot and never a dull day but, truth be told, I do love a challenge! Read on for a glimpse into our serials metadata standardisation journey.

An introduction

If you've never heard of the BFI (British Film Institute) let me introduce you. We are a charity founded in 1933 and governed by a Royal Charter (1) to:

- encourage the development of the arts of film, television and the moving image
- promote their use as a record of contemporary life and manners
- promote education about film, television and the moving image generally, and their impact on society
- promote access to and appreciation of the widest possible range of British and world cinema
- provide funding for film across the UK
- establish, care for and develop collections reflecting the moving image history and heritage of the United Kingdom

Collections fall under two distinct areas. The Archive (2) is responsible for moving image, stills, posters and unpublished papers. The Library (3) for published print and electronic materials.

The Library

The library started in 1934 as "a small pile of books" (4) and has over time developed into a highly respected non-lending research library, now free to all.

Our collections encompass various aspects of the moving image and its development and are international in scope with comprehensive coverage of the moving image in Britain. The following illustrates the library's current collection size and scope.

Books – 45,000 titles

- You may be surprised to know that our earliest book, *Ars Magna Lucis et umbrae* by Athanasius Kircher, was published in 1646, a date long before the invention of cinema. *The webcam as an emerging cinematic medium* by Paula Albuquerque, published in 2018, is one of our most recent acquisitions. Monographic series are treated as books.

Journals – 5,242 titles

- Our earliest journal, *The Optical Magic Lantern Journal and Photographic Enlarger*, first published in June 1889 is also our longest running title continuing to this day as *Screen International*.

Pamphlets – 7,000 items

- A pamphlet to us means an item of less than 80 pages (or quite flimsy), mostly books but also the occasional annual.

Annuals – 2,000 titles

- Annuals include annual reports, directories, yearbooks, filmographies, statistical reports and similar.

Festivals catalogues

- The festivals collection consists of folders containing any number of published items relating to a film or television festival. I cannot at present give you a number of items but I can tell you that it occupies around 120 linear metres of shelving.

Distribution catalogues

- The distribution and sales catalogues of 8 or 16mm film and other formats, all 20 linear metres of them, are effectively uncatalogued. We also hold lantern slide and apparatus catalogues from the 19th and early 20th century.

Press cuttings files – 4.1 million

- We no longer buy newspapers but our digitised archive of press cuttings is accessible as PDFs on-site via our catalogue (Collections Search).

The Serials collection

In 2011, as a direct result of major restructuring, the library's collections of Periodicals, Annuals and Festivals were brought together to form the Serials collection and the role of Serials & e-Resources Librarian created. The Distribution and sales catalogues were included a few years later.

The Serials team consists of myself and an Assistant Librarian, with the assistance of Reader Services Assistants whenever possible. A year on from the restructuring we moved to a new site, migrated our data to a new LMS, and started looking into RDA. We are still keenly feeling the effects of these actions!

Advanced search	
Search in	Library catalogue - All
Title - exact	Moving image catalogue - All
Word(s) from the title	Film and Television Works
Title of Film/TV work	Film and Television Manifestations
Periodical title	Film and Television Archive Items
Author	Library catalogue - All
Corporate author / Publisher	Books
Name keyword	Serials
Subject keyword	Articles
	Press Cuttings
	Special Collections catalogue - All
	Scripts, Documents and Ephemera
	Posters and Designs
	Stills
	Persons and Institutions
	Workflow

Figure 1 – Collections Search - Advanced search - Serials

Necessity and opportunity

Faced with the task of managing what had been disparate collections, each with their own cataloguing history, we have spent the last few years trying to bring cohesion to serials cataloguing practice. Our priorities, in terms of aiding discovery and improving metadata quality, have been informed by necessity, otherwise known as opportunity.

It is fair to say the resulting actions involve a fair amount of data cleaning and data enhancement peppered with hints of work arounds.

1. Following relocation of the collection and migration to a new LMS (Adlib) our first priority has, by necessity, focused on all serials records showing, at the very least, accurate holdings and location information at title level. We are sadly some way off achieving this goal as much of the information is incomplete, inaccurate or in the wrong fields, and some is only visible to staff in the client version.
2. Our second priority has been to take the opportunity to standardise, as far as possible, cataloguing practices for all our serials and enhance the metadata, regardless of whether the information is visible to our users or not at this stage.
3. Our third priority is to help our users discover the collection and delve deeper. We no longer index our journals to the extent that we did before the restructuring so we have taken the opportunity to assign subject headings at title level, starting with our current journal titles which number around 230. Only five thousand or so left to go!

Standardisation

All serials have been, and continue to be, catalogued as a hierarchy with individual records at title, issue and article level. This was our common starting point across the serials collections. We have traditionally used minimal level AACR2 as it was sufficient for our needs and, in case you're wondering, we still catalogue from scratch and have never used MARC.

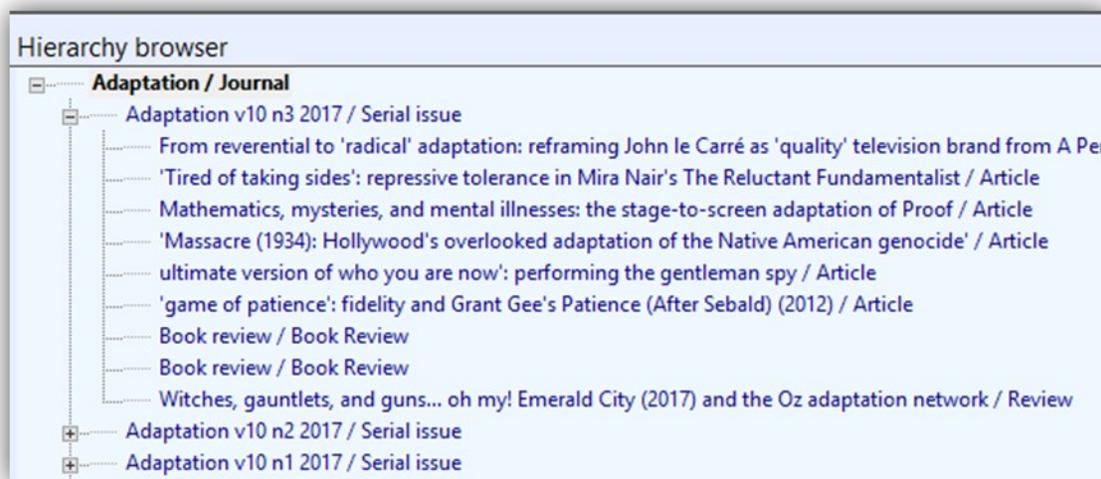


Figure 2 - Journal record showing hierarchy as visible on Adlib client

Title	Adaptation
Material	Journal
Language	English
Notes	Abstracting & Indexing: British Film Institute (BFI); International Federation of Film Archives (FIAP); Film & Television Literature Index (EBSCO).
Publication place	Oxford, Great Britain
Publisher	Oxford University Press
ISSN	1755-0637

Site	Status	Loan category	Holding summary
BFI Southbank - Closed Access Stack	Available	Reference Only	v1 n1 2008 - v10 n3 2017

Figure 3 – Journal record as viewed on Collections Search, our online catalogue

For better or worse we decided on the following, as a start. Our decisions were shaped by consideration for local custom and practice, new developments in library metadata standards, and a lack of resources and suitable formal training.

Title

- Journal titles previously capitalised all words except articles, i.e. *Sight & Sound* instead of *Sight & sound*, whereas Annuals did not. Our current titles have been brought in line with RDA guidelines however we still follow a custom style for the numbering of serial issues.
- At title level Help and Alternative titles are added to facilitate exact searching. A Help title could be the title excluding the subtitle – this helps eliminate all issue titles from the search in client - or the title with the word replacing the symbol, so ‘and’ instead of ‘&’.
- At issue level Alternative titles are recorded to allow searching under an issue’s given title. For example, the annual *IRIS plus* publishes subject specific titles for each issue. The alternative title for IRIS plus n4 2014 is *The influence of new technologies on copyright*. This isn’t visible to the user but can be searched for within the client.

Author

- Editors and corporate authors have traditionally only been recorded for annuals. These will now also be added for other serials records as required. At title level changes in editor over time are noted in a notes field. At issue level the editor’s name remains an authorised access point. This isn’t visible to the user but can be searched for within the client.

Publication details

- Place of publication now includes both place and country because searching by country remains important to our users. Traditionally annuals recorded city as place of publication, journals recorded country and festivals the location of the event.

Title	Berlin International Film Festival Berlinale Internationale Filmfestspiele Berlin European Film Market		
Material	Festival		
Language	German, English		
Holding	Collection includes main catalogue and, where available, supplementary publications such as the catalogue of the European Film Market. Please ask library staff for further details regarding exact holdings.		
Publication place	Berlin, Germany		
Publisher	Internationale Filmfestspiele Berlin		
Subject keyword	Festivals: film ; Catalogues: film ; Berlin ; Germany		
Site	Status	Loan category	Holding summary
BFI Conservation Centre - Multi Media Vault - Festivals	Available	Reference Only	(1st) 1951 - (67th) 2017 Incomplete.
BFI Southbank - Library Office	Available	Reference Only	(68th) 2018

Figure 4 - Example of a Festivals record

Identifier

- ISSN is recorded at title and issue level. Where one exists we also record ISBN at issue level. This isn't visible to the user but can be searched for within the client.

Subject headings and Classification

- All serials will eventually have subject headings assigned but shelfmarks are assigned to annuals and distribution catalogues only as these, once catalogued, are filed in the same sequence as the books. Journals are filed by title. Festivals filed by year, place and date.

Holdings

- Holdings information, divided by physical location, is recorded in the copies database. This applies to both title and issue level. No such information is recorded for articles and chapters.
- For serials included in the Serials Management Module (these are journals mostly) we are gradually adding holdings start and end dates.
- Annuals records, on the whole, currently show holdings in the abstract field but they don't always show location, or much other information for that matter, and where they do it's likely to be inaccurate. Their records will eventually match the rest of the collection.

Title	A.M.P.A.S.:Screen achievements record
Material	Annual
Publisher	AMPAS

Figure 5 - Annual with minimal information, as viewed on Collections Search

Title	L' Agence du Court-Metrage catalogue			
Material	Annual			
Abstract	also called "Regie TV-Cable catalogue". Vol 1(1986), Vol 2('88), Vol 3('90), Vol 4('93), Vol 5('95), vol.6 ('97), vol 8 (CD Rom), vol.9 (2002)			
Publication place	Paris			
Publisher	L'Agence de court-metrage			
Subject keyword	Catalogues: film ; France			
Shelfmark	Site	Status	Loan category	Holding summary
791.411:944	BFI Southbank - Closed Access Stack	Available	Reference Only	

Figure 6 – Example of an Annuals record

Other changes

- We now add dimensions for all serials. This is visible in the client only.
- We add accompanying material information. This is visible in the client only.
- We have started creating separate records for digital versions.

Title	16mm catalogue		
Material	Pamphlet		
Corporate author	Anglo EMI Film Distributors		
Pagination	100 p.		
Notes	No date.		
Shelfmark	Site	Status	Loan category
	BFI Conservation Centre - Multi Media Vault	Available	Reference Only

Figure 7 - Example of a distribution catalogue record

Conclusions

The above tentative steps towards a standardised approach to cataloguing serials are certainly not set in stone and we are a long way off what I would consider to be ideal for our data. What I can say is that I am very excited by the recent consideration given to serials by the IFLA LRM, PRESSoo and RDA. In my own mind I have certainly freed our data from the confines of a library record and think in terms of a web of entities and relationships, of data that is useful and connected beyond the restrictions of a library management system.

What I am less certain of is how we bring the required changes about, however, there was talk of just such a revolution at the CIG18 conference. I am very pleased to be a member of a group determined to actively promote awareness and appreciation of metadata professionals and their continually evolving role in this field.

Biography

Following her graduation Anastasia embarked on a long career at the British Film Institute and is currently enjoying the latest challenge of serial and electronic resource management. Her qualifications include the BA (Hons) Librarianship & Information Studies (PNL, 1990), the BA in Hispanic Studies (Birkbeck, 1995) and the MSc in Information Science (UCL, 2013). She is a current member of the CIG committee.

Bibliography

1. **British Film Institute.** The BFI's Royal Charter. *BFI*. [Online] 2018. [Cited: 08 12 2018.] <https://www.bfi.org.uk/about-bfi/bfi-s-royal-charter>.
2. **British Film Insitute.** BFI National Archive. *BFI*. [Online] 2018. [Cited: 08 12 2018.] <https://www.bfi.org.uk/archive-collections>.
3. **British Film Institute.** BFI Reuben Library. *BFI*. [Online] 2018. [Cited: 08 12 2018.] <https://www.bfi.org.uk/education-research/bfi-reuben-library>.
4. **Butler, Ivan.** *To encourage the art of film: the story of the British Film Institute*. London : Robert Hale, 1971.

Summary

The Jisc Data Community Group have made a number of recommendations to Jisc to tackle problems with bibliographic metadata quality and supply in the context of aggregating library metadata in the National Bibliographic Knowledgebase (NBK).

Introduction

Jisc are developing a national-scale shared service - the National Bibliographic Knowledgebase (NBK) - that will build upon and surpass the functionality of [Copac](#)³ eventually including catalogue data from more than 225 academic and specialist libraries. By doing so, it will more effectively support the management of library collections so that they are optimised for contemporary research and learning needs.⁴ The NBK will underpin plans to create a UKRR⁵ for monographs and potentially the development of a knowledgebase for the management of e-resources.

In parallel with NBK development Jisc convened funded community groups to look at issues and possible interventions that might be made to enhance the quality and efficiency of library bibliographic and holdings data.⁶ This article summarises the report and recommendations from the Data Community Group, which considers data quality and standardisation to be the primary factors in ensuring the success of the NBK.

The aim of the group was to create a framework to map different levels of metadata against activities that can be supported by each. This framework would provide a practical understanding of metadata standards across the sector, and the practical effects of metadata quality and aggregated quantity on user experience.

The methodology involved contributing questions to a survey written with the other community group projects. The survey was circulated nationally in June 2018, which generated a response from 99 institutions.

1. Jisc is the "UK higher, further education and skills sectors' not-for-profit organisation for digital services and solutions" <https://www.jisc.ac.uk/> [accessed 23rd November 2018]

2. Data Community Group members: Nick Barratt, Senate House Library, University of London, Paul Cunnea, National Library of Scotland, Jane Daniels, Cardiff Metropolitan University, Clare Hudson, London School of Economics and Political Science, Vanessa Lacey, Cambridge University, Diana Massam, Jisc, Thomas Meehan, University College London, David Morgan, Royal Holloway University of London, Amy Staniforth, Aberystwyth University.

3. Copac is a union catalogue that exposes rare and unique research material by bringing together the catalogues of over 100 major UK and Irish libraries, see <https://copac.jisc.ac.uk/about/> [accessed 9th November 2018]

4. See the beta NBK as of November 2018: <https://www.jisc.ac.uk/national-bibliographic-knowledgebase> [accessed 9th November 2018] and video <https://www.jisc.ac.uk/rd/projects/national-bibliographic-knowledgebase#> [accessed 16th November 2018]

5. UKRR Research Reserve, see <https://www.ukrr.ac.uk/> [accessed 23rd November 2018]

6. See the CCM tools blog from the NBK <http://blog.ccm.copac.jisc.ac.uk/> [accessed 20th November 2018]

An initial analysis of the survey responses was collated by the Data Community Group and presented at the Cataloguing and Index group conference (CIG18) in Edinburgh September 2018⁷ and Jisc have published a wider summary of the survey.⁸

Aim 1: to develop a greater understanding of the matching algorithms currently used by OCLC and Jisc to underpin the NBK.

Conclusions:

The existing clustering and deduplication process in CBS (the OCLC system which the NBK is using) is very complex and is supported by a set of technical system documentation developed for internal use.

The Data Community Group feels that it is important that an explanatory document is developed and circulated so that users can gain a better understanding of the process and rules.

In addition, the guiding principle for choosing a base or master record is that there should not be a preference for records from any particular library. The choice of base record has yet to be decided by OCLC and again, the group feels that user input via Jisc is important in determining the choice of base record.

Once a base record has been chosen, then additional useful fields from other consolidated records can be added. Jisc and OCLC are working on how to consolidate a subset of records for a group of libraries, and what impact this would have on the overall quality of the consolidated record. Further updates on this work will be issued by the NBK team in due course.⁹

The NBK team have also been working on the raw data with elastic search, with a 'quality' indicator based on the data in fields and number of fields in particular records. The results of this will also be circulated to the community for further discussion.

Recommendations:

1. Jisc and OCLC share an explanatory document which details the process for deduplication of data in CBS.
2. Jisc – in consultation with users – determine the choice of base record.
3. The community be notified when these pieces of work are in a position to be shared, with ongoing input by the community into the way the base record is identified and enhanced for matching or import functionality.
4. Proactively involve the community in the evolution of algorithms, using the Community Groups' survey to understand why users have failed to engage with development or application opportunities to date. With more widespread, targeted and effective input, the community can advocate NBK as a strategic metadata collaboration.

7. See CIG18 slideshare: <https://www.slideshare.net/CILIPCIG/national-bibliographic-knowledgebase-survey-data-quality-subgroup-initial-survey-findings> [accessed 16th November 2018]

8. Jisc, Headline Summary <https://blog.ccm.copac.jisc.ac.uk/2018/11/20/community-data-groups-survey-headlines/> [accessed 20th November 2018]

9. The Data Group have recently responded to a consultation on record selection in the OCLC CBS system (the database currently used for NBK data), for example.

5. Roadmap development for planned updates to matching algorithms, with regular communication and explanatory narrative about what's been changed and why, to encourage user engagement. This will help the community drive the direction of future change, placing the library sector as the commissioner for the NBK and associated toolkits, with Jisc as the contractor and agent of delivery.

6. In the short term, we recommend a webinar on 'elastic search' functionality and opportunity of use, to sit alongside other tools in development. This will promote community activity around catalogue enhancement, and allow contributors to prioritise metadata improvement tasks and resubmit better quality records.

7. In the medium term, guidance on matching and searching algorithms should be produced so that it is clear when institutions should create new records, for example, and this good practice can be adopted as institutional policy nationally.

8. Furthermore record vendors will need to ensure that records for large data sets e.g. eBooks, are compiled to internationally recognised standards to obviate the need for extensive normalisation at import.

Aim 2: to map how metadata standards are applied across SCONUL and RLUK libraries, allowing metadata quality thresholds to be defined more clearly

Conclusions:

There is a growing awareness that the metadata that libraries utilise for local acquisitions, cataloguing & discovery workflows is of variable quality. Libraries are required to invest thousands of pounds each year on additional cataloguing work to upgrade deficient metadata. This has resulted in discrepancies and variant cataloguing practice across the sector, exacerbating the problems facing any composite knowledgebase.

There is more confidence in supplier data for print material than e-book, perhaps a product of the way data is processed (with item by item print cataloguing generating greater confidence in the accompanying data compared to bulk uploads or database provision for e-books). Only 50% of libraries report errors back to suppliers, and within this group there is a higher proportion of reporting for purchased records. This is an urgent issue for the library sector to address with data suppliers, given the amount of time spent on checking and amending metadata.

Data quality problems appear to be particularly acute in relation to large records sets for e-resources. Checking is often abandoned altogether with concomitant impact on search & discovery. The group feel that the library sector needs to regain control over data creation, sharing & maintenance. This would produce more efficient local and national cataloguing workflows, freeing up time for data improvement and original cataloguing projects.

A part of the problem may lie in the vendors assuming that they are supplying good quality metadata, particularly if libraries do not report errors. An aim of the NBK could be to negotiate the provision of national agreed minimum standard records for all formats which libraries could use and enhance locally, or purchase enhancements from suppliers.

Recommendation:

9. The Group recommends that Jisc work with BIC (Book Industry Communication), NAG (the National Acquisitions Group), publishers, and systems, content and record vendors to develop a new commercial business model that places library client requirements at the heart of all transactions. At the moment libraries spend additional money on resource to check data, and then further resource enhancing it to a required standard (which may then differ from institution to institution).

10. Instead, the group recommends that NBK becomes a central data resource, a 'trusted partner' for all newly created e and print metadata, ingesting an agreed 'base level' from suppliers that libraries can then ingest into their system once a resource has been purchased. Jisc would then act as a vendor-neutral knowledge base supplying data on demand, with transparent standards that link to bibliographic fields and matching algorithms. Users can then choose to add content to their records should they wish, either using their own resource or purchased from the original data supplier, and then share these enhancements as contributors to the NBK.

11. Equally, there needs to be a greater emphasis placed on controlled and standardised vocabulary when cataloguing or improving metadata. The creation of the above model for future data would free up experienced resource to tackle issues around legacy data, standardisation and raising quality – potentially at consortia level, given the increasing alignment of work within groups such as WHELP, SCURL, G4W, Mercia, M25 and White Rose. Unleashing a 'hive mind' of professional cataloguers, freed up from replicated manual checking across multiple institutions, would save resource and improve data, and develop our uses of it in the web environment.

Aim 3: to review current practice amongst Copac users, with wider review of other methods of data sharing and systems or data management practice

Conclusions:

There is a concern that institutions are unable to act on Copac error reports to improve the quality of data. If NBK is to work effectively and ensure errors are reported and corrections uploaded to improve the quality of data, there needs to be an incentive for users to engage with NBK.

It would be useful to understand Copac's view of the data they receive, particularly common issues. NBK algorithms and benchmarks around data quality and standards need to be clear, given the expectation that NBK will become the 'definitive record'.

The group felt that other upload habits should be understood at a more granular level too (for example the English Short Title Catalogue), to ensure that NBK has captured all sources of information.

Recommendations:

12. Further surveys of the sector are required to understand how and when uploads are made, and how the data is used. This might form a series of single-question micro surveys to obtain richer, more granular data about the community. Understanding who responds to the surveys might be as important as the information they provide, so that efforts can be made to involve other institutions who have not engaged.

13. For complex legacy records, it would be useful for NBK to develop the elastic search feature, and also to create standard brief level rules—where Library Management Systems (LMS) support these—for assessing metadata quality.

Aim 4: to scope a potential wider review of historic cataloguing practice which would provide an understanding of required future standardisation work.

Methodology:

It was intended that Aim 4 might follow from some of the survey results. However, due to timing issues this was not possible. However, it is clear from views expressed at various conferences, seminars and workshops held whilst the project was undertaken that there is a longer-term interest in this work – in particular from the Collection Management: Share the Experience group and National Acquisitions Group.

Conclusion:

It has not been possible to draw any concrete conclusions at this stage, other than noting that there is interest in pursuing further work.

Recommendations:

14. Commission a separate project to collate historic cataloguing practice, using national or regional consortia such as SCURL, WHELP, Mercia, M25, G4W and White Rose, potentially framed by the Collection Management: Share the Experience group. From a professional standards perspective, it would be useful to involve CILIP and the CILIP Cataloguing and Index Group (CIG) in conversations.

15. As in Aim 3, a series of short pulse surveys is recommended to continue the conversation across the sector and build up a greater understanding of historic practice. This would form the basis of retrospective work, particularly if future data creation can be standardised as outlined in Aim 2.

16. The anonymised raw data from the survey should be made openly available to encourage further research across the sector.

Conclusions

The exercise has raised awareness of some wider structural and economic challenges facing the sector that need to be addressed if NBK is to realise its full potential.

The group recommends a 'state of the nation' conversation, mediated by Jisc, between representative library bodies and national groups involved with data (SCONUL, RLUK, regional consortia, CILIP and ARL (Association of Research Libraries) in the US, BIC and NAG), data providers (for example Elsevier, Ebsco, ProQuest), record vendors (for example BDS), and potentially LMS providers (for example Innovative, ExLibris, OCLC, SirsiDynix, Capita).

The aim would be to focus on what libraries actually require from bibliographic records, metadata and open access publications to create a new hybrid commercial model based on library need. The alternative is to bypass the data suppliers and use national libraries, consortia and specialist staff to create value-added content for future linked data. Libraries are businesses and need to look at operational efficiency. The current market set up does not work economically, nor are libraries obtaining what they need to function properly.

The existing consortia have a key role to play in both developing the new economic model, the development of a vendor-neutral knowledgebase for the management of e-resources, and forming a regional strategy for moving beyond NBK to UKRR-Monograph with storage solution. Jisc should convene an initial meeting with the consortia to explore possible collaborative metadata projects and an alignment of codes of practice and governance.



A library sector research programme needs to be developed in parallel with the above, with funding for students to undertake infrastructure work. This could be linked to ‘library school’ dissertations with funding from CIG, RLUK, AHRC and potentially Arts Council grants. There should also be a review of the training, qualifications and career pathways for colleagues creating, sharing, enriching and maintaining metadata to ensure a robust and resilient workforce now and in the future.

Part of the research programme would explore the relationship with non-LMS catalogue platforms that are being used to discover or surface content, such as Google Scholar. There is a concern that unmediated content coupled with an absence of information literacy skills is eroding the ability of researchers to locate and understand content.

Libraries need to define their role in the information economy more clearly as trusted purveyors of quality data, and build an information literacy skills environment that forms one of the earliest parts of the learning journey in higher education. This will require advocacy from across the sector but particularly from library Directors to their University executive groups.

Biography

The NBK Data Community Group is made up of librarians from across the UK Higher Education sector. In our various roles we are responsible for library metadata and metadata practices, and we see the National Bibliographic Knowledgebase as an opportunity for libraries to reshape the metadata landscape in the UK.

At Conference this lightning session followed a preview of the results of a survey distributed to higher education and special libraries by Jisc to inform the development of the [National Bibliographic Knowledgebase](#)¹ which will replace [COPAC](#)² a national merged catalogue of the holdings of over 100 UK libraries. These include national, university and special libraries.

Dr Amy Staniforth, a member of the [NBK Community Data Group \(Metadata\)](#)³ shared some of the preliminary findings on data quality issues and cataloguing practices. As a co-member of this Group it occurred to me that there might be an additional use for the NBK. What if it could also include catalogue records for resources that had been lost or withdrawn and were considered lost to the national collections?

What is metadata memory?

The term “Metadata Memory” was actually coined by Diana Massam (Jisc Project Manager - Copac Collections Management Tools) after my initial approach to Jisc with this suggested additional function for the NBK. I.e. that of metadata archive which would:

- Preserve bib records for re-use even in the absence of library holdings e.g. last copy withdrawn/lost/destroyed
- Allow records to bear witness to what *had* existed thus providing a valuable resource for researchers

Surveying the Cataloguing Community

I was asked by Jisc to do some more work on the idea and to assess whether there was library support for it. In September 2017 I sent out an email explaining the proposal and asking some basic questions.

Text of email:

“Like many people I have been thinking about the fantastic opportunities that the National Bibliographic Knowledgebase provides us with. In short, a national collections discovery service par excellence fuelled by the creation, sharing and curation of excellent metadata.

As cataloguers we know the value of good metadata for discovery. We aim to describe what is extant in our collections but when we withdraw the last copy, either as part of a collection management policy or through loss, our records are suppressed or deleted. They cannot be re-used by other cataloguers or bear witness to the existence of that resource in scholarly research.

1. National Bibliographic Knowledgebase. Available at: <https://www.jisc.ac.uk/rd/projects/national-bibliographic-knowledgebase> [Accessed 24/11/2018]

2. COPAC. Available at: <https://copac.jisc.ac.uk/> [Accessed 24/11/2018]

3. arratt, Nick. (2018) NBK Metadata Mapping Project. Available at: <https://blog.ccm.copac.jisc.ac.uk/2018/05/22/nbk-metadata-mapping-project/> [Accessed 24/11/2018]

Many of us work with materials that are not published so are not catalogued by our Copyright Deposit Libraries for inclusion in the British National Bibliography. Perhaps we should be preserving more metadata for re-use and equally importantly for use in future research?

*What is your initial reaction to this idea?
Should /could it be done?*

Perhaps you have a local policy on archiving records or contribute records to other catalogues safe in the knowledge that your metadata will be preserved?"

Responses

There were 7 responses, 6 from the HE sector and 1 from a library consultant. Not a vast number but really interesting feedback/suggestions and most importantly everyone was supportive of what I was proposing.

Here are the answers to the questions.

- Should it be done – YES
- Could it be done? – YES. One of the respondents offered this advice: “The Record could continue to exist in COPAC ... [NBK] ... [with] a prominent statement along the lines of 'No longer any UK holdings' where the holdings are normally displayed.” At the moment only records with library holdings are retained in COPAC, so once the records are deleted from the service there is no record of those resources ever having existed unless they are included in the BNB or retained as suppressed records in the owning library’s catalogue.
- Institutional metadata retention policy or export for preservation? – Sometimes.

Based on my small sample there was a division between:

- Those who retain suppressed records permanently e.g. to facilitate usage analysis. I didn’t have time to include this in the session in Edinburgh but one of the respondents mentioned the workflow that they followed as a Capita customer. Capita Base (which used to be Talisbase) is a reciprocal customer database of millions of bibliographical records. When Capita libraries withdraw stock the bibliographical records are suppressed locally and the customer holdings updated on Capita Base.
- Those who retain them until external agencies such as COPAC or OCLC have deleted their library’s holdings from those aggregated services.
- No policy at all.

I produced a summary of the findings for the [Jisc Collection Management Community Advisory Board meeting held on 17th October 2017⁴](#) and there was support for investigating metadata memory further.

My purpose in presenting at CIG18 was to disseminate the idea and my initial research and gauge the level of interest amongst the cataloguing community for this extension of the NBK functionality. Metadata Memory ties in well with the collection management service that Jisc hope to provide using the data in the NBK but currently the priority has to be securing the contribution of as many library holdings as possible so that we can all reap the benefits of this fantastic project.

4. Massam, Diana. (2017) Report of the CMCAB October meeting. Available at: <https://blog.ccm.copac.jisc.ac.uk/2017/12/19/cmcab-october-meeting/> [Accessed 24/11/2018]

Please do think about sending your holdings to Jisc if you have not already agreed to contribute. The additional exposure of your resources and the ability to benchmark your collections using [CCM Tools](#)⁵ are just a couple of the benefits that will accrue. But what additional, more specific, benefits would metadata memory secure?

Cataloguing efficiencies

Given the costs of cataloguing it makes sense to preserve records so that:

- Descriptive and subject cataloguing is preserved in perpetuity. Something else to consider here is the maintenance of metadata. There are currently no plans to provide maintenance for access points (Names/Subjects) in the NBK but as the service matures perhaps this is something that could be investigated as it represents another value-added service for NBK contributors?
- There is re-use, not recreation, of records for resources lost, destroyed or withdrawn but subsequently re-discovered.
- It supplements the work done by our copyright deposit libraries. The records that they produce live on for posterity in the [British National Bibliography](#)⁶

NBK as Research Resource

What about the benefits to users? This really gets to the heart of what we do as cataloguers and the inestimable value of our work. By cataloguing resources we affirm their existence & make them discoverable. If we preserve the metadata, when the last physical copy is withdrawn or a digital file is deleted, then the records act as book marks in history for those resources.

Retaining records for resources that are no longer accessible:

- Adds to the history of scholarly publishing
- Helps researchers track the rise, decline and evolution of ideas and subjects

Next Steps

I anticipate re-surveying the cataloguing community in 2019, perhaps as part of the NBK Phase 2 activities and I would like to present the findings at the fifth *Jisc Share the Experience : Collection Management* event being held at Royal Holloway, University of London 11th-12th June 2019. Please do complete the survey when it arrives next year.

Metadata Memory would be a service for cataloguers by cataloguers and represents a chance to reclaim our metadata and forge strong collaborative workflows for the creation, sharing, enrichment, maintenance and preservation of records. In the meantime please do contact me with your comments, suggestions and questions.

Biography

Jane Daniels is Bibliographical Librarian at Cardiff Metropolitan University. In addition to cataloguing in Higher Education Jane has been a Subject Librarian; part of a technical services team in a public library authority; and a local government information officer. She started her career working in medical libraries. She is a graduate of the College of Librarianship Wales in Aberystwyth.

jdaniels@cardiffmet.ac.uk @JaneDaniels1

5. Copac Collection Management Tools. Available at: <https://ccm.copac.jisc.ac.uk/> [Accessed 24/11/2018]

6. British National Bibliography. Available at: <http://bnb.bl.uk/> [Accessed 24/11/2018]

The last decade has seen a dynamic shift in the way libraries think about catalogues and bibliographic data. Machine Readable Cataloguing (MARC), which was once a staple in the cataloguing environment, is now seen as tired and nearing obsolescence. In order to move beyond MARC, forward-thinkers in the library community have tasked themselves with finding a potential replacement. Linked Data is thought to be a contender for a replacement. While it may still be a few years down the road before Linked Data is fully adopted by the greater community, we can begin now to decide to prepare for such a transition.

Backstage Library Works (Backstage) is on a quest to begin helping libraries prepare for the eventual need to shift catalogues from a MARC environment to a Linked Data environment. While Backstage is not currently involved with *transformation* of library data into linked data, we are providing options to help prepare for that transformation through our Resource Description and Access (RDA) Enrichment and Uniform Resource Identifier (URI) Enrichment services.

Linked Open Data (LOD) in and of itself is not a new concept but the use of Linked Open Data within a library environment has been discussed since 2007.¹ LOD is a way to share information on a much more global scale through the semantic web.² In order for data to be shared it needs to be housed in a format that can be readily used and open for data mining.

BIBFRAME and other linked data frameworks for libraries use RDF (Resource Description Framework) as a way of packaging the metadata for eventual use on the semantic web. URIs are heavily used within a LOD environment in order to connect the data appropriately and accurately across the web. Before one can begin to talk about where these URIs come from, one must take one more step back and talk about library data as it currently exists and how it can be fully exploited.

Library's bibliographic data largely resides in the MARC format and is usually expressed either through the Anglo-American Cataloging Rules 2 (AACR2) or RDA standard. AACR2 was developed in the late 1970s and in 2004 RDA began being developed by the RDA Joint Steering Committee as a replacement for AACR2.³ The overarching goal of RDA was to begin recording data in a way that would be either transferrable to a non-MARC format or be catalogueable in other metadata formats to give greater ability for discoverability through searching utilities.⁴

Even though RDA was fully adopted by the Library of Congress in 2013, with much training and experimentation for years before that, not all institutions have made the jump. Some are still originally cataloguing new materials in AACR2 while accepting copy records in either AACR2 or RDA.

1. RDA Steering Committee. (2018). *How does RDA support linked data?* Retrieved from RDA Frequently Asked Questions: http://www.rda-rsc.org/content/rda_faq#13

2. (2018). Retrieved from Linked Data: <http://linkeddata.org/>

3. Joint Steering Committee. (2015). *Background*. Retrieved from Joint Steering Committee for Development of RDA: <http://www.rda-jsc.org/archivedsite/rda.html#background>

4. Joint Steering Committee. (2015). *Scope*. Retrieved from Joint Steering Committee for Development of RDA: <http://www.rda-jsc.org/archivedsite/rda.html#scope>

Others have also begun all cataloguing in RDA but haven't been able to attempt to convert the remainder of the catalogue to RDA because doing so manually would take significant time and money. With the push to move away from a MARC environment it is becoming ever more important to begin thinking about bringing older library data into RDA .

The additional elements available within RDA are meant to assist with increased findability by faceting data that were previously coded in unsearchable MARC fields. These new fields have the potential to be indexed and searched within a MARC environment. Much of the faceted terminology within RDA are also controlled by various registries and term lists, meaning that any institution using that registry for terminology will all be using the same vocabulary for their metadata. The use of these standard terms lists also leaves room for the possibility of future automated control of these terms as well as linking in a LOD environment using URIs.

Whether or not the Library Management System (LMS) can make use of the new RDA elements or the faceted nature of the elements is inconsistent. But the desire of institutions to have their data fully accessible by their users will hopefully continue to push the LMS vendors to be more receptive to changes.

In 2011 Backstage Library Works participated as an RDA test partner with the Library of Congress and also began developing automated workflows to convert AACR2 data into hybrid data; data that is technically still AACR2 but has been enriched with RDA elements to bring more consistency across the catalogue. Options for automated RDA Enrichment are extensive and can be customised according to any institution's preference. No institution catalogues in the same way so automated enrichment needs to be flexible to accommodate specific requests.

There are numerous changes from AACR2 to RDA so just a few of the changes that can be automated will be highlighted now. One of the major changes from AACR2 to RDA was the decision to expand abbreviations in certain descriptive fields. Appendix B in the RDA toolkit highlights which terms in which languages should be expanded to the fullest form of the term.⁵ The majority of these can be seen in the MARC 26X, 300, 504, and relator terms while other terms such as "Department" are to be spelled out everywhere.

MARC Fields: Before

```
300 ____ > a118 p. > billus. > c25 cm.
504 ____ > a"Footnotes and bibliographies": p. [103]-118.
700 1_ > aBaird, Joseph Armstrong. > eed.
```

MARC Fields: After

```
300 ____ > a118 pages : > billustrations ; > c25 cm
504 ____ > a"Footnotes and bibliographies": pages [103]-118.
700 1_ > aBaird, Joseph Armstrong, > eeditor.
```

5. RDA Steering Committee. (2017, Apr 11). *Appendix B*. Retrieved from RDA Toolkit: <http://access.rdatoolkit.org/rdaappb.html>

Additionally, the previously used MARC 260 publication information field has now been changed to the MARC 264 field. The 264 also includes new MARC indicators that give a greater ability to describe what the information within the 264 is, such as publisher versus manufacturer versus copyright. There has been debate whether a 260 field can, or should, be changed through automated means into a 264 field and for a few years the conversion was *not* recommended by the Program for Cooperative Cataloging (PCC). Backstage, however, has worked tirelessly to create a conversion that has a high level of confidence. Even in 2018, OCLC has been working on the 260 to 264 conversion for their RDA coded records.

MARC Fields: Before

```
260 ____ > aBerlin :> bDistanz ;> aBerlin :> bGestalten, distributor, > cc2015.
```

MARC Fields: After

```
264 __ 1 > aBerlin :> bDistanz, > c[2015]
264 __ 2 > aBerlin :> bGestalten, distributor
264 __ 4 > c©2015
```

An area of perceived contention comes with the MARC 245\$h field, the General Material Designation (GMD). In RDA, the GMD is converted to various Content, Media, Carrier (CMC) fields for fuller description. However, some LMS cannot utilize the CMC fields fully so many institutions are continuing to opt to retain the GMD within the 245 as well as having the CMCs added.

MARC Fields: Before

```
007 ____ vfrcbahou
008 ____ 891027s1988 nju120f|||| |v|leng d
035 ____ > a (IaU) 001034707IOW01
035 ____ > a (OCoLC) 20763154
035 9_ > aAEK3490
040 ____ > aIaU, cIaU
100 1_ > aRumelhart, David E.
245 10 > aParallel distributed processing, h[videorecording] /> cDavid Rumelhart.
300 ____ > a1 videocassette (120 min.) :> bsd. :> c1/2 in +> eguide.
```

MARC Fields: After

```
007 ____ vf cbahou
008 ____ 891027s1988 nju120 f vleng d
035 ____ > a (IaU) 001034707IOW01
035 ____ > a (OCoLC) 20763154
035 9_ > aAEK3490
040 ____ > aIaU, cIaU, dUOrBLW
100 1_ > aRumelhart, David E.
245 10 > aParallel distributed processing /> cDavid Rumelhart.
300 ____ > a1 videocassette (120 min.) :> bsound ;> c1/2 in +> eguide.
336 ____ > atwo-dimensional moving image > btdi, 2rdacontent
337 ____ > avideo, bv, 2rdamedia
338 ____ > avideocassette, bv1, 2rdacarrier
340 ____ > apolyester, b1/2 in.
344 ____ > bmagnetic, 2rda
346 ____ > aVHS, 2rda
```

As can be seen above, audio visual materials see even more RDA fields being added to accommodate other types of information about the piece. Printed music also sees some additions as can be seen in the following example.

MARC Fields: Before

```
100 1_ > aAssad, Sérgio, > ecomposer.
240 10 > aSuite brasileira, > mguitar, > nno. 3
245 10 > aSuite brasileira no. 3 : > bguitare seule = solo guitar / > cSérgio Assad.
264 _1 > aSaint-Romuald : > bDoberman-Yppan, > c2015.
264 _4 > c©2015.
300 ___ > a1 score (19 pages) : > c31 cm
500 ___ > aFingerings by Thomas Viloteau.
```

MARC Fields: After

```
100 1_ > aAssad, Sérgio, > ecomposer.
240 10 > aSuite brasileira, > mguitar, > nno. 3
245 10 > aSuite brasileira no. 3 : > bguitare seule = solo guitar / > cSérgio Assad.
246 31 > aSolo guitar
264 _1 > aSaint-Romuald : > bDoberman-Yppan, > c2015.
264 _4 > c©2015
300 ___ > a1 score (19 pages) : > c31 cm
336 ___ > anotated music, > bntm, > 2rdacontent
337 ___ > aunmediated, > bn, > 2rdamedia
338 ___ > avolume, > bnc, > 2rdacarrier
348 ___ > ascore, > 2rda
382 ___ > aguitar, > n1, > s1
383 ___ > ano. 3
```

The last area to highlight today are RDA rules 6.2.2.4 and 6.11.1.4 regarding expressions with more than one language. Together the rules state that each language should be listed but it is not explicit as to if they should be in a single \$l. However, to be properly controlled headings, the languages should be split into their own access points. As such, an option may be selected to separate these access points accordingly.

MARC Fields: Before

```
240 10 > aIsla de los cánticos. > lEnglish & Spanish
```

MARC Fields: After

```
700 12 > aVaz Ferreira, María Eugenia, > d1875-1924. > tIsla de los cánticos. > lEnglish.
700 12 > aVaz Ferreira, María Eugenia, > d1875-1924. > tIsla de los cánticos. > lSpanish.
```

RDA rules and standards are dynamic and are updated regularly. Whatever automated methods are being used for conversion from AACR2 to hybrid records, the process needs to remain flexible to be able to incorporate changes easily and quickly.

RDA conversion of a catalogue creates a base on which to further prepare the data for a linked data environment. As previously mentioned, some of the new RDA fields include terminology that is controlled by various term lists. These term lists have been issued with linked data in mind and have URIs associated with the individual terms.

When we talk about URIs within library data there are two different types of URIs. The first is a URI that is associated with a description of the entity within the MARC field, which I shall continue to refer to as just URI. The second is associated with the Thing represented within the MARC field, a Real World Object (RWO).

While there are third-party utilities available to libraries to have URIs and RWOs added to bibliographic data, there is an argument to having an authority control vendor add this information for access points. Authority Control vendors often track additional forms of the bibliographic heading that are not currently tracked either in the authority record or various term lists. These variant forms may include common typographical errors or even older authorized forms of the heading or term, again not all of which appear as a cross reference. So, an authority control vendor will often have a higher matching rate for adding URIs than using another utility.

It is during the process for term matching just described that the URI or RWO links are retrieved and inserted into the Bibliographic heading. Most often the URIs are pulled from regularly maintained authority files, such as the Library of Congress Name and Subject Authority Files or term lists from the RDA Registry. RWOs are often retrieved from the Virtual International Authority File or the International Standard Name Identifier, though there are many other possible sources available. In order to determine which sources should be expressed as authority URIs in a \$0 and those which should be RWO URIs in a \$1, the PCC Task Group on URIs in MARC has issued a document that helps identify and build both types of URIs entitled *Formulating and Obtaining URIs in MARC: A Guide to Commonly Used Vocabularies and Reference Sources*.⁶

As mentioned already, other term lists such as the RDA Registry has URIs available for other terms. These URIs are available for CMCs as well as other MARC fields such as the 34X MARC fields for audiovisual materials. Additionally, other MARC fields such as the 043 geographic designation code as well as the 050 Library of Congress Classification number may also have URIs added. Additional fields are also being reviewed and added that have the possibility of URI linkage.

```

043  __ ▼ae-hu--▼0http://id.loc.gov/vocabulary/geographicAreas/e-hu
050  _4 ▼aM2149.2.P8▼bP75 2015
      ▼0http://id.loc.gov/authorities/classification/M
336  __ ▼anotated music▼bntm▼2rdacontent
      ▼0http://rdaregistry.info/termList/RDAContentType/1010
337  __ ▼aunmediated▼bn▼2rdamedia
      ▼0http://rdaregistry.info/termList/RDAMediaType/1007
338  __ ▼avolume▼bnc▼2rdacarrier
      ▼0http://rdaregistry.info/termList/RDACarrierType/1049
600  10 ▼aOláh, Miklós,▼d1493-1568.▼0http://id.loc.gov/authorities/names/n88049265
700  1_ ▼aSzoliva, Gábor,▼eeditor.▼0http://id.loc.gov/authorities/names/no2015165623
830  _0 ▼aBavarica et Hungarica ;▼v2.▼x1789-5375
      ▼0http://id.loc.gov/authorities/names/no2009105604

```

The question arises whether it makes more sense for URIs to be added while still in the MARC environment or if they should be added as one is transforming data into LOD. There are strong arguments within the library community, and those at the forefront of linked data exploration, that URIs should be added in the MARC environment.

6. Program for Cooperative Cataloging Task Group on URIs in MARC. (2018, Feb 14). *PCC Task Group on URIs in MARC*. Retrieved from Library of Congress: https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf

By doing so, the transformation of data into Linked Open Data should be smoother as the URIs will already be present, thus reducing the amount of manual reconciliation. Also, it gives libraries a chance to start thinking about how Linked Open Data will work for them and begin experimentation.

We also should not ignore our non-MARC data. Some collections have been stored in a format other than MARC, perhaps an XML schema for a digital collection. If we want a true linked data environment for all our collections, we need to consider adding URIs to this data for eventual transformation into an RDF linked data world as well.

```
<image>
  <image_title>Lister Institute Group Photograph with names</image_title>
  <image_iap_image_no>L0068570</image_iap_image_no>
  <image_keywords_unauth>
    <_>Institute;</_>
    <_>International ACM SIGGROUP Conference on Supporting Group Work;</_>
    <_>ACM SIGCHI International Conference on Supporting Group Work;</_>
    <_>members;</_>
    <_>fellows;</_>
  </image_keywords_unauth>
</image>
<uris>
  <uri_lc_name>http://id.loc.gov/authorities/names/no98079700</uri_lc_name>
  <uri_viaf>http://viaf.org/viaf/312890028</uri_viaf>
  <uri_lc_name>http://id.loc.gov/authorities/names/no2009138784</uri_lc_name>
  <uri_viaf>http://viaf.org/viaf/160752016</uri_viaf>
</uris>
```

Regardless of where your institution currently stands regarding RDA and Linked Open Data, it's not too late (or too early) to begin thinking about the future needs of your users and to begin preparing your data for a world where collections are much more accessible. Backstage is committed to helping libraries of all sizes try to meet those needs and will continue to modify our automated processes to do so. We're all working towards a common goal of increased access, so let us all continue to collaborate to make that a reality.

References

- (2018). Retrieved from Linked Data: <http://linkeddata.org/>
- Joint Steering Committee. (2015). *Background*. Retrieved from Joint Steering Committee for Development of RDA: <http://www.rda-jsc.org/archivesite/rda.html#background>
- Joint Steering Committee. (2015). *Scope*. Retrieved from Joint Steering Committee for Development of RDA: <http://www.rda-jsc.org/archivesite/rda.html#scope>
- Program for Cooperative Cataloging Task Group on URIs in MARC. (2018, Feb 14). *PCC Task Group on URIs in MARC*. Retrieved from Library of Congress: https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf
- RDA Steering Committee. (2017, Apr 11). *Appendix B*. Retrieved from RDA Toolkit: <http://access.rdatoolkit.org/rdaappb.html>
- RDA Steering Committee. (2018). *How does RDA support linked data?* Retrieved from RDA Frequently Asked Questions: http://www.rda-rsc.org/content/rda_faq#1

Biography

Casey Cheney has been with Backstage Library Works for 16 years and spent 10 years designing custom solutions to metadata needs with libraries of all sizes. She resides in Provo, Utah with her trusty feline sidekick, Reggie, and enjoys travelling wherever the opportunities take her.
ccheney@bslw.com

From 5th - 7th September 2018, people who work with library metadata gathered from across the UK and further afield for the Cataloguing and Indexing Group conference in Edinburgh. I am very grateful to CIG for the bursary which enabled me to attend. I hoped to go as cataloguing has been a significant element of my work in libraries but I wanted to learn more about current issues, projects, problems and initiatives in the wider library and information profession.

There were several papers on ethical issues in metadata, perhaps reflecting CILIP’s ‘Big Ethics Conversation’ and new ethical framework. Issue 191 of *Catalogue and Index* demonstrated that the world of cataloguing and metadata is certainly not exempt from ethical consideration. Jo Maxwell spoke about the active development work being done by the DDC Forum to update the classification system which has been criticised variously for its Anglo-American bias, treatment of indigenous peoples, women, and religions apart from Christianity. Certain additions, such as the ‘vegan cooking’ heading have literary warrant, whilst others are developed by crowdsourcing, community engagement and collaboration. Gordon Dunsire spoke on the second day of the conference about ethical issues in cataloguing using the case study of gender in name authority records.

A highlight of the conference for me was Joshua Barton’s paper on the ethics of zine cataloguing at Michigan State University. As subcultural publications, zines do not necessarily belong in libraries. They defy publication norms and are ultimately incompatible with our assumptions about library metadata; that the creator should be identifiable and associated with a work, and that items should be discoverable. Zines are made for a community that promotes a network of support and care, but as institutions which increase visibility of zines through metadata, libraries can augment the creators’ vulnerability. Cataloguers’ insistence on overcoming the often purposeful obfuscation of the creator’s identity can be dangerous for the privacy of personal information from families, employers and authorities. Joshua drew on Caswell and Cifor’s concept of radical empathy in archival work as well as Fox and Swickard’s ideas to suggest that we should ask for the consent of subjects included in collections before outing identities, as well as consulting and collaborating with subject communities in our descriptive work more widely. This challenged my assumption that more detailed metadata is paramount in improving the experience of library users; nuance is needed to also consider zine creators’ rights and privacy.

Improving discoverability, however, remained an overriding theme of the conference, with access to collections an important principle of creating, sharing and enriching metadata. Robert Miles talked about the [Endangered Archives Programme](#) run by the British Library, and their aim to enrich the metadata and searching capacities of its new website which hosts digitally preserved archives that would otherwise be vulnerable to neglect and physical deterioration. Improvements such as adding creation dates rather than the date of digitisation, content type facets and subject and place authorities have improved consistency and discoverability of these important and fascinating archival documents.

On the second day of the conference, Amy Staniforth talked about a project at Aberystwyth University to improve discoverability of library resources and mark the contribution that metadata makes to student experience. Amy explained how small, bespoke metadata enhancements, for example of subject headings, contents and summary notes, can enhance student satisfaction through improved discoverability of resources not on undergraduate reading lists. Amy used the case study of improving catalogue records of the 1847 ‘*Report into the state of education in Wales*’. The volumes are more commonly known as the ‘*Blue books*’ but were un-findable by this name. She suggested that the most significant improvements can be made to rare, historic material, which, once made more accessible through accurate and comprehensive metadata, are more likely to be found and used by students and recommended by their supervisors.

There was some discussion afterwards about how to measure the success of a metadata enhancement project. This in turn raised more interesting points about how metadata is displayed to the user and the efficiency of libraries' search and discovery facilities.

There was a variety of papers which demonstrated the wide array of metadata projects across the world. In the first paper of the conference, Jane Daniels presented the Wales Higher Education Libraries Forum's (WHELFL) achievements in collaborative cataloguing in Alma. Andrea De Cornò talked about cataloguing special collections in RDA with a case study of the London Library's *Per Nozze* collection of printed ephemera, addressing specific challenges such as new syntax for pagination and foliation and RDA relator terms which are useful but not consistently adequate for this special collection. James Mitchell, meanwhile, talked about cataloguing artworks and three-dimensional sculptures held by the National Library of Scotland explaining the use of hybrid standards and multiple controlled vocabularies: Library of Congress Subject Headings as well as Getty Art and Architecture Thesaurus and Library of Congress Thesaurus of Graphic Materials genre terms.

Helen Williams gave a lightning talk on the London School of Economics's website improvement project, recommending that metadata specialists use their expertise beyond a library setting but to be prepared for different styles of working! Helen also gave a talk later in the afternoon on archiving the LSE's blog outputs. The blogs are an important and significant research output for the university, for example being the most-tweeted material by political scientists, but were hosted in a volatile web context. Helen talked through specifics of how it was attempted to automate parts of the process of archiving, as well as some recommendations and tips for planning and implementing projects more generally.

Annelyn Lim and Rothman Lim's paper outlined the impressive work they have done in a project to create Fi-SAF, a database of subject authorities for Filipiniana materials. The huge project which the two have undertaken was a sharp reminder for the audience of the importance of uniformity.

Penny Doulgeris' presentation on how the International Atomic Energy Agency migrated from a (pretty nightmarish-sounding) non-MARC, non-integrated library management system to a new cloud-based LMS. Penny discussed how they mapped existing fields in their records to MARC21 standards in order to migrate the metadata, and ended by remarking that the nature of her job has changed drastically since the system change!

Sally Rimmer explored the process of integrating the University of Derby's institutional repository into the library catalogue using MarcEdit: a tool I would love to learn to use after several of the speakers praised its utility and especially Will Peaden's lightning talk about its 'magic' for improving metadata at Aston University.

There were other sessions that supported more practical knowledge and tools such as Paul Shackleton's interactive session 'Everything you wanted to know about WorldCat'. WorldCat is a resource I have only ever known from a user's perspective so it was fascinating to hear more about the work that goes into the catalogue's deduplicated master record model with sophisticated matching process, as well as the different routes libraries can take to upload their records. Louise Howlett's session helped me to understand the history and tricky intricacies of serial cataloguing; "*you make statements about the past and assumptions about the future*". Louise gave an update on the development of PRESSoo (apparently not an acronym) and the influence it has had upon the Library Reference Model. In another interesting update, Amy Staniforth reported on the findings of a survey on current cataloguing practice and concerns in the context of Jisc's new National Bibliographic Knowledgebase (NBK) which will soon replace Copac.

I found it really interesting to hear about the process of research where accurate metadata was a critical resource in Melissa Terras' keynote on 'Finding professors in children's literature'. Melissa spoke about the library catalogues she used in the search for illustrated representations of academics in children's literature: WorldCat, the British Library, National Library of Scotland, National Library of Australia.

Melissa also consulted digital libraries: the Hathi Trust, Google Books, Digital Public Libraries of America, as well as specialist children's literature collections such as the Baldwin Library of Historical Children's Literature and the International Youth Library, but explained the difficulties of searching for *illustrated* representations since they are not mentioned in standard library metadata. Moreover, Melissa also highlighted the various sources of unofficial metadata held on sites such as eBay, AbeBooks, GoodReads, LibraryThing and BookDigests which had a very different style of description to libraries and could offer leads that could then be followed up on the library catalogues. The entire project, Melissa explained, began on Twitter, and unofficial channels of research such as Twitter and [Tumblr](#) were also invaluable for crowdsourcing examples.

Melissa's two books, [Picture-Book Professors: Academia and Children's Literature](#) and [The Professor in Children's Literature: An Anthology](#), both recently published in open access format, document the 328 academics found in 289 books. Each example was consulted in a physical copy of the book although many of the titles were freely available in full text online. Books published before 1900 in the public domain, which had been digitised with full-text searching options made research for that era easier, although there were in fact many more examples over the 20th century when more universities were established and the number of juvenile literature books published increased.

Melissa's research notes that in children's literature, academia is represented as people, not the university as place. There were many examples of the professor as archetypal wise, old man but also as clueless, silly and baffled professors who do nothing useful. Tropes such as the 'evil' professor appear only in publications after World War Two; public fear of science after the result of nuclear bombs permeated children's literature and popular fiction. But the first example of a woman academic with a name comes in a book published in 1992, 142 years after the first named male professor appears. Furthermore, women with academic power are not likeable, popular characters and thus, Terras asks, what are we teaching our children about knowledge, and power, and who is allowed to hold it? At the start of her paper, Melissa confessed she was nervous to be speaking in a room of library professionals but I felt it was a really valuable element of the conference to have an academic share the process of research and 'working out' in humanities scholarship. Her plea to fix the persistence of book reviews being 'more relevant' than the book itself according to library catalogue searches was met with knowing nods.

Ultimately, the conference made me feel part of a wider network of people involved in cataloguing and indexing and it was inspiring to see the range of projects and research the CIG community is involved in. Thanks again to the CIG Committee for the generous bursary that enabled me to attend. If you want to catch up then you can see many of the [presentations online](#) and browse tweets from the conference under the hashtag #CIG18.



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

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

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

Karen Pierce: PierceKF@Cardiff.ac.uk

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

Book reviews: Please contact the editors.

ISSN 2399-9667

CIG website: cilip.org.uk/cig

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

CIG twitter: @CILIPCIG

Tags from the CIG blog on specific areas of interest:

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