ISSN 2399-9667

# Catalogue and Index

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

## Editorial

June 2019, Issue 195

Welcome to C&I 195, an issue all about metadata standards, with a particular focus on RDA. RDA stands, of course, for Resource Description and Access*, and, as the successor to AACR2, it is intended to provide a means for those involved in bibliographic metadata (and beyond) to describe materials in a standardised, interoperable, and useful way. RDA has divided opinion since its introduction in 2010, but it's safe to say that it is here to stay, and we thought now would be a good time to look back over the past few years of RDA implementation, and also look forward to what the future will bring for the standard. In an increasingly interlinked metadata world, however, it's also important to be aware of standards from outside the remit of traditional cataloguing, and we hope that you find the articles on other standards as enlightening as we did.

On the RDA side, Alan Danskin shares a revised version of his presentation from #CIG18, offering a helpful overview of the conceptual changes to come in RDA 3R, the broader implications that these will have for our metadata in general, and the specific approaches being adopted by the British Library to make the most of these changes.

Thanks to Jenny Wright you can also read about how the European RDA Interest Group (EURIG) is contributing towards the new RDA toolkit, and get entity-based updates on the future of Bibframe and a Hungarian FRBRisation project.

---

*It also stands for the Research Data Alliance and the Riding for the Disabled Association, amongst others, and according to Wikipedia is also the name of one of John Dee's minor Enochian angels, but we sadly won't have room in this issue to explore all of these.

CILIP SPECIAL INTEREST GROUP

OPEN ACCESS

## Contents

We have practitioner account of RDA implementation from Charlotte Smith, who was involved in the introduction of RDA to the University of Cambridge libraries. Her article highlights both the unique challenges of working at such an idiosyncratic institution, and some of the universalities of training, such as the tension between providing subject-specific or more generalist material.

Moving beyond RDA, we have a fascinating article by Paul Walk on the Dublin Core Metadata Initiative, custodians of the most widely-known of the many vocabularies that are key to the future of metadata. It's an important reminder of the potential fragility of the internet's metadata infrastructure, but encouraging to see DCMI's approach to tackling this.

Graham Bell provides an ever-valuable insight into metadata standards used outside – but ever so close to – the librarian's usual experience, with an overview of EDitEUR. EDitEUR are the creators and custodians of ONIX and *thema,* standards used throughout the publishing world.

Alan Danskin expands upon his article above with an outline of the BL's Collection Metadata Strategy, and provides us with an insight into how they intend to implement changes to their metadata in practice, without sacrificing conceptual integrity. No small task…

For those working with rare materials, Iris O'Brien provides an overview of the efforts to bring DCRM in line with RDA, and the Rare Books and Manuscript Section Bibliographic Standards Committee's current recommendations for cataloguers.

Will Peadon reflects on JISC's 22nd May 'Metadata Day' event, gives us his thoughts on the future of metadata, and asks the big questions about the role of MARC in encoding and exchanging data.
We finish with reviews of *Ethical Questions in Name Authority Control* and *Social Tagging in a Linked Data Environment,* and our letters page. As always, we greatly appreciate your responses and feedback, so do please get in touch!

This issue we also welcome Philip Keates as our new co-editor.

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

Philip Keates: philip.keates@kcl.ac.uk

# Implementing RDA 3R, or Why are we implementing RDA again?

Alan Danskin, British Library

This is an updated and revised version of the presentation given at #CIG18 in Edinburgh.

## RDA Restructure and Redesign Project

RDA 3R is a project initiated by the RDA Steering Committee and Co-publishers, with the approval of the RDA Board. The project has the objective of restructuring and redesigning the RDA Toolkit. RDA needs to change to be brought into compliance with the IFLA Library Reference Model (LRM); the underlying RDA data needs to be reconfigured to improve flexibility and efficiency of support for updating and translation.

## Business Case

RDA 3R offers the British Library the potential to address some of the strategic challenges we face in managing collection metadata as identified in, *Unlocking the value*, and developed in our new strategy document, *Foundations for the future*, which takes the strategy through to 2023. LRM and RDA are central to our proposals for an (internal) Target Metadata Model that will enable convergence of collection metadata standards and break down the internal silos that are a barrier to productivity and discovery.

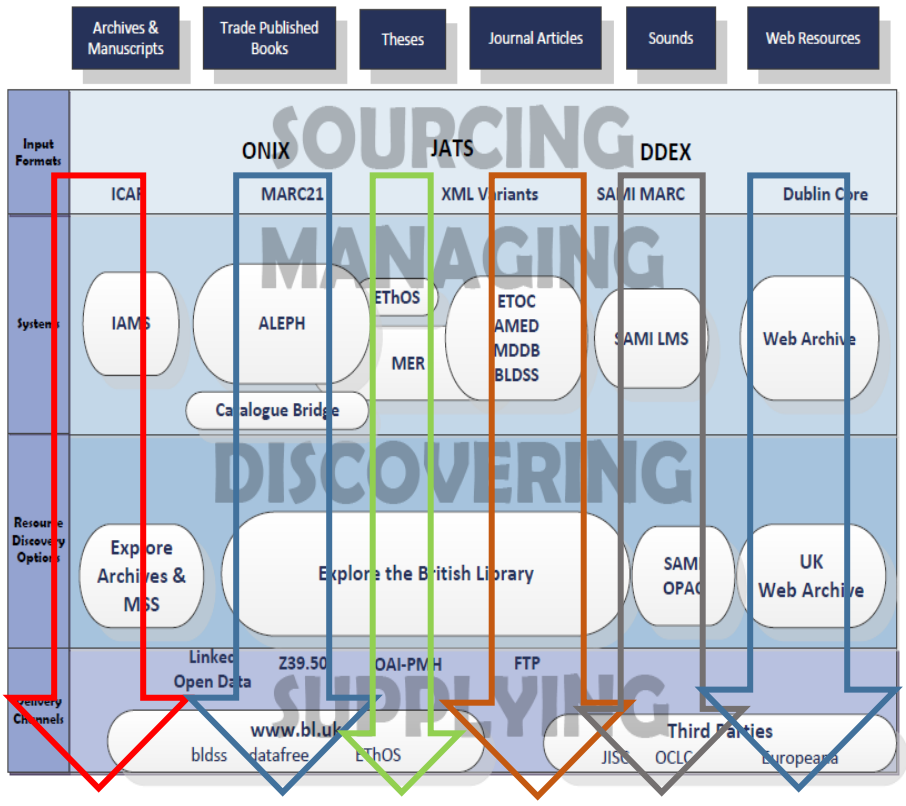


**Fig. 1: Metadata Silos**

The Target Metadata Model complements the Library Systems Transformation (LST) programme to replace the Digital Library System and our main cataloguing systems: Aleph, IAMS (Integrated Archives and Manuscript System) and SAMI (Sound and Moving Image) system.

LST is a phased programme and it will be several years before we replace Aleph. In the interim, we won't be in a position to derive all the benefits of LRM or 3R, so our aim is to minimise the impact on our users, staff, and other stakeholders. As in 2013 we will work with our stakeholders, including Legal Deposit Libraries and CIP agent and the users of our services. Where possible we will share documents and training. And we will maintain services throughout the transition. Before discussing implementation in more depth, I want to look briefly at some features of 3R.

**New Entities**

LRM introduces new Entities: Nomen, Timespan and Place. This enrichment of the model makes RDA more hospitable to time-based media, including sound recordings and facilitates convergence with metadata created in accordance with other cataloguing traditions or by other cultural heritage sectors.

**Recording Methods and data provenance**

RDA now offers four different recording methods for each element, where applicable:
1. Unstructured description
2. Structured descriptions
3. Identifier
4. IRI (internationalized Resource Identifier)

The utility of the metadata recorded increases with each level. For example an unstructured description can only be reliably indexed as keywords, whereas an IRI is a persistent linked data compliant actionable link which returns information about an entity or controlled term.

RDA also includes provision to record the provenance of metadata. This is increasingly important for us because we need to automate processes by which metadata from external sources can be reused without compromising quality.

**Elements**

**Manifestation: manifestation statement**
Manifestation statement is a new manifestation element that enables data to be transcribed directly from the source without having to parse it into separate sub-elements. In the short term, these changes should enable our early printed materials cataloguers to use RDA rather than DCRM for description of the manifestation. In the longer term, we would expect to reduce the time cataloguers spend on transcription of information by extending the use of tools, such as light pens or scanners. MARC 21 does not currently support the manifestation statement.

**Labelling of relationship designators**
Relationship designators are a type of element. They are distinguished from other elements by their properties. All elements have a "domain" (the entity or class to which they belong) but relationship designators also have a "range". The range is the entity that is the value of the relationship, i.e. the entity to which the relationship points. RSC has expanded the existing relationship designator labels to make these underlying properties more explicit. For example, **author** becomes **author agent**, which is refined to **author person**, **author family**, **author collective agent**, **author corporate body.** These labels are not intended, nor are they suitable, for public displays.

Where the semantics of the relationships are expressed through the MARC 21 encoding: the status quo may be an option for MARC users
e.g.

X00 1= + $e author = author person
X00 3= + $e author = author family
X10 + $e author = author corporate body

It is uncertain whether a value of "collective author" may be necessary to distinguish collective agents that are neither families nor corporate bodies, but capable of authorship.
Another option would be to record the IRI for the relationship in subfield $4.

**Guidance Chapter**

**Aggregates**
Aggregates are defined as manifestations which embody two or more expressions. The new guidance is the result of research carried out by the RDA Working Group on Aggregates into an aspect of the FRBR model that was poorly understood. Although most manifestations are aggregates this will not be bibliographically significant in most instances. RDA provides models for the different types of aggregates and guidance on how to record them.

**Diachronic Works**
Diachronic works are works that are planned to be embodied over time rather than in a single act of publication. Serials and integrating works are examples of diachronic works. Because the contents of different manifestations (e.g. print and digital) often diverge the concept of the WEMI lock has been introduced. This means that a change to specific attributes of the manifestation or expression will result in a new work. In practice this is unlikely to have a significant impact, but the ISSN is now considered to be an identifier for the work not the manifestation. The ISSN-L identifies a cluster of related serials.

**Fictitious and non-human appellations**
Current guidance in relation to fictitious and non-human entities has also been changed. A fictitious entity, e.g. Geronimo Stilton, is assumed to be a pseudonym of a person or other Agent, even if the true identity is unknown. A non-human entity (e.g. an animal) that is associated with a statement of responsibility is treated as an entity that is external to RDA. This is because the LRM definition of person is limited to human beings.

**Implementation**

The British Library's internal implementation is being planned in four broad phases.

**Timeline**

There are still unknowns; most significantly we don't know when the 3R project will be declared finished. Throughout the period we will work with CIG and CILIP to keep the community informed and we will engage with stakeholders, including the Legal Deposit Libraries and BDS, the Agent for the Cataloguing in Publication programme.

As you can see (on the next page), we are presently in phase 3. The RSC has announced the stabilization of the English text. Stabilization means that the English text cannot be changed in a way that would impact translations or outcomes; corrections may still be made and errors should still be reported. Stabilization gives cataloguing agencies the confidence to proceed with the preparation of translations and local documentation. Local documentation includes application profiles; policy statements and workflows.

**Timeline**

| | Phase 1 | Phase 2 | Phase 3 | Phase 4 |
|---|---|---|---|---|
| **Date** | June-September 2018 | October-March 2018 - 2019 | April 2019-2020? | 2019/2020? |
| **3R** | **Beta Testing**<br>September Release | RSC Meeting Ottawa<br>RSC Finalise English Text<br>Develop Load Scripts for Policies | **April: English Text Stabilised**<br>Policy documents loaded<br><br>Translation(s) loaded<br>December RDA Board<br><br>MARC RDA Working Group | **RDA 3R Signoff**<br><br>**Current toolkit switched off 2021?** |
| **British Library** | **Evaluation**<br>Review Beta Version<br><br>Revise RDA/MARC mapping<br><br>Stakeholder Engagement | **Preparation (1)**<br>Assess Outcomes<br>Test Load Scripts<br>Application Profile<br><br>Update policies<br><br>Stakeholder Engagement | **Preparation (2)**<br>Write new Policies<br><br>Revise Workflows<br><br>Develop training<br><br>Stakeholder Engagement | **Implementation**<br>Deliver Training<br>• CIP Agent<br>• BL Cataloguers<br><br>Distribute 3RData<br><br>Stakeholder Engagement |

9

**Fig. 2 BL Implementation Planning Phases**

**Application profiles**

Application profiles specify the entities, elements and vocabulary encoding schemes that are expected by an application that is going to use the data. The specification includes whether an element is repeatable or non-repeatable; whether it is mandatory or optional and may include other information, including the vocabulary encoding scheme or string encoding schemes to be used in relation to specific values.

It is expected that application profiles will be developed by RSC or national cataloguing agencies rather than individual institutions; although institutions may wish to adapt these profiles to their own needs.

EURIG (European RDA Interest Group) has developed a general application profile based on the current profile used by DACH (German Speaking Community) as a proof of concept. EURIG submitted the profile to RSC in March and RSC decided to establish a working group to further develop the work.

**Policy Statements**

The British Library and other national agencies publish their policy statements in the RDA Toolkit. In the new toolkit it will be possible to “subscribe” (i.e. select) specific institutions’ policy statements to view. Policy statements will appear in the right hand content pane, in line with the instruction to which they apply.

The British Library has revised the text of existing policy statements to conform to changes in RDA terminology. However we have to write policy statements for instructions relating to new elements. Our statements are being used to test the script for loading the policy statements, which also matches them to the relevant instruction. This is proceeding more slowly than we expected, hence the lack of policy statements in the beta toolkit.

**Workflows**
The British Library developed and maintains several detailed workflows for the current toolkit. Workflows provide a "narrative" to aid navigation of the instructions. They are an important training aid, but also continue to be used by experienced cataloguers. Our workflows have also been adopted and adapted by other institutions.

We have concluded that the structural changes to RDA increase the need for a narrative to mediate the toolkit contents. As the existing workflows cannot be easily adapted we are undertaking a comprehensive redesign. We are basing a test workflow on the content of the EURIG application profile. The workflow will provide guidance on getting started and some general points to remember. We aim to have fewer workflows than currently exist and for them to be shorter in order to reduce maintenance.

Entries will be provided for elements in the application profile and for other elements that we consider significant. We aim to avoid repetition of RDA instructions but will focus on guiding cataloguers efficiently between elements. Where possible we will include examples encoded in MARC 21. We will also link out to other standards and relevant documents.


Recording the title of a manifestation¶
Take the title of the manifestation from the appropriate source of information¶
Follow the instructions under Manifestation: title of manifestation for prerecording and recording¶
Record the title of manifestation as a title proper. (Note, that by doing so you are also satisfying the instruction to record a value of Nomen)¶
If the title is in non-Roman script, transliterate the title following the appropriate transliteration table¶

**Fig. 3: Example entry from draft workflow**

**Impact**
So far as we are able to assess at this time, 3R implemented within your current environment will be essentially neutral, i.e. have no long term impact on productivity.

I don't expect 3R to have a significant impact on catalogue users, so let's look at the impact on, cataloguers, their managers and system managers.

**Cataloguers**
From a cataloguer perspective the first challenge will be understanding what has changed in the Toolkit. The navigation and presentation of information is completely different, so cataloguers will need training and time to get to grips with it. The numbering has been removed and perhaps the hardest thing to get used to will be the absence of any imposed order or sequence of elements. Hence the priority we are giving to workflow development.

There is a lot of new terminology. In practice many of the new entities and associated attributes may not be used differently, but cataloguers will have to understand how the terminology relates to what they are doing. The biggest hurdles will probably be nomen string; manifestation statement, aggregates and diachronic works. We are developing training materials to familiarise staff with these concepts.

As previously described, there are now 4 recording methods for most elements. To some extent our choice may be guided by MARC; for example 245 $c is an unstructured statement about responsibility for a work; whereas 700 is an authorised access point. In other cases, a policy may be necessary to provide guidance. The old distinctions between options will have gone. Instead, each element will be associated with a number of optional instructions. Some choices will depend on the kind of material being catalogued or the situation; other options will require policy guidance.

We are promised many more examples and richer examples. Cataloguers will need to be aware of the different types of examples and how to use them.

**Cataloguing Managers**
The new administration interface has only just been released. At present I have only been able to access the Manage institutional view, which enables selection of institutional settings for language, primary policy statements, and display options for examples and element reference. I have not been able to access the Manage Users and Profiles module, it is not clear whether reporting is included in this module or will be a separate function.

The main impact of 3R will be on those responsible for managing cataloguing departments or teams. Much of the documentation you already have will be obsolete and will need to be revised. New documents will be required to reflect new elements and the many more options. The extent to which you have to develop or amend your own documentation will vary; where possible you should seek to reuse documentation. Where local applications of policy vary from BL/LDL you can create an annotated bookmark.

I don't expect training to be as intensive as it was in 2013, but it is important that cataloguers receive adequate orientation to enable them to use the toolkit and apply RDA with confidence.

**System Managers**
In the short term, the changes to your existing system will be for the most part familiar activities. Do the new entities have an impact on indexing? MARC changes (see next section) will have a knock-on effect on templates, labels and configuration. Depending on how you set up templates, you may want them to match RDA changes to element names. If you use Macros these may need to be edited or new ones created to reflect changes in terminology.

Internal or external validation tools will need to be updated. If your data is transformed for export to other systems, such as a discovery system, the transforms may need to be updated. Citation software can also be affected. Data you import from third parties or export to third parties may need to be tested and profiles amended.

**MARC 21**
The British Library will continue to distribute RDA data in MARC 21 for the foreseeable future. I don't expect 3R records to be flagged differently from current RDA records.

In MARC 21 a 3R record looks very much like a current RDA record. As discussed above there is an issue regarding relationship designators.

```
LDR   nam a2200265 i 4500
008    180208r20182017enk 000|f|eng|d
020     $a 9781473654891 $q paperback
020     $z 9781473654907 $q ebook
020     $z 97814736665644 $q digital download
040     $a Uk $b eng $c Uk $e rda
041     $aeng $h it
082 04 $a853.92 $2 23
100 1   $a Varvello, Elena, $e author.
240 10 $a Vita felice. $l English
245 10 $a Can you hear me? / $c Elena Varvello ; translated by Alex Valente.
264  1 $a London : $b Two Roads, $c 2018.
300    $a 260 pages ; $d 20 cm
336    $a text $2 rdacontent
337    $a unmediated $2 rdamedia
338    $a volume $2 rdacarrier
500    $a Translated from the Italian.
500    $a Originally published: 2017.
586    $a English PEN Award
651  0 $a Italy $v Fiction
655  0 $a Suspense fiction. $2 gsafd
700 1  $a Valente, Alex, $e translator.
```

**Fig. 4: Mock-up of a MARC 21 record created using 3R**

Following on from the presentation given by Thurstan Young at ALA Annual in 2018, the RDA/MARC Working Group is being reconvened by Library of Congress. The group will consider the changes to MARC to accommodate new RDA elements, recording methods, etc. In the event that new MARC fields are introduced to support aspects of 3R there will be full notification of these changes in advance. So far as we are aware, there are no dependencies on MARC for implementation of 3R. The earliest any proposals could be submitted is ALA Annual 2020. Changes that are agreed will require system configuration changes and may affect indexing, templates, macros, validation and transformations, both within your system and in associated systems or applications.

**Training**

Implementation will include orientation and training for cataloguers and other RDA users. The core modules will cover:

- IFLA Library Reference Model (LRM), which integrates and supersedes the earlier FRBR, FRAD and FRSAD models
- Using the RDA Toolkit
- RDA Terminology and British Library policy, practice and documentation
- How to catalogue with RDA Toolkit
- Recording Methods

A number of additional modules are also likely to be necessary to cover specific topics or types of resource. For example:

- Authority Control
- Aggregates
- Diachronic Works: Serials, Works in Progress and Integrating Resources

We expect that the hands-on training required will be significantly less than was required for implementation in 2013. Nevertheless, there will be an impact on productivity, for which we are developing mitigation strategies. We also plan to update our popular RDA in a Day course for delivery in 2020 and ALA has also announced a series of webinars on different aspects of 3R.

**Concluding remarks**

Following signoff of the 3R Project, there will be a 12 month window for implementation, during which the current and new toolkits will run in parallel. At the end of that period the current toolkit will be switched off.

There are too many dependencies for the British Library to give a firm date for implementation. It is most likely to be in the second half of 2020, which is later than we initially anticipated. We have discussed coordinating our implementation with Library of Congress, and other national libraries, as we did in 2013. As in 2013, we will also be coordinating with the CIP agent (BDS) and the Legal Deposit Libraries to ensure a smooth transition.

Any short term benefits are likely to be dependent on MARC developments, but in the longer term the structural changes introduced by LRM and 3R will increase the range of materials that can be catalogued using RDA and enable more efficient processing.

**Links**

*IFLA Library Reference Model (LRM)* *https://www.ifla.org/publications/node/11412*

*RDA Toolkit (Beta site)* *https://beta.rdatoolkit.org/*

*Foundations for the future: the British Library's Collection Metadata Strategy, 2019-2023*

*https://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2019-2023.pdf*

*RDA and MARC 21: the impact of the 3R Project*. Gordon Dunsire, James Hennelly and Thurstan Young

Presentation http://gordondunsire.com/pubs/pres/M21AndRDA.pptx

Slideshare https://www.slideshare.net/GordonDunsire/m21-and-rda

*Getting a handle on the new RDA Toolkit*: https://www.rdatoolkit.org/index.php/news

RDA Toolkit Youtube channel:

General https://www.youtube.com/c/RDAToolkitVideo

BetaSite Toolkit training https://www.youtube.com/playlist?list=PL1Gb2VwmOhgTlQdOabM-xHWtlfSVea5Z6

RSC Presentations: http://www.rda-rsc.org/rscpresentations

A range of presentations covering different aspects of RDA 3R

## RDA in Europe

**Jenny Wright, Chief Metadata Officer, BDS**

The European RDA Interest Group, **EURIG**, has a representative on the RDA Steering Committee, and is provided with feedback from the UK metadata community by members from the British Library, National Library of Scotland, and CILIP. Institutions interested in EURIG can find information on the group and membership here http://www.rda-rsc.org/europe

The UK Committee on RDA provides a forum for discussing issues within the UK, and supporting users and prospective implementers of the RDA standard. Anyone interested in joining the UK Committee on RDA should get in touch with the chair Jenny Wright jenny.wright@bdslive.com

A significant element of EURIG's contribution to the RSC is to be mindful of the heterogenous nature of libraries. European libraries represent many different languages, and some non-Latin scripts. In the coming year, EURIG is agreeing a revised text for the collaboration with RSC; providing feedback on new Toolkit to the RSC; working on application profiles for the new Toolkit and organising a satellite conference in Greece, prior to IFLA's WLIC2019, on August 21.

The annual meeting of EURIG was held in Budapest in May 2019. The focus of the meeting was to discuss the new RDA Toolkit, specifically its usability, and what is required from application profiles. In addition there were several interesting presentations, many of which are shared on the EURIG website http://www.rda-rsc.org/europe/meetings/2019meeting

Examples are not officially part of RDA, but an examples editor has been appointed and a Working Group is anticipated. EURIG plans to contribute, as the lack of relevant examples is a significant issue for European users. In addition, a new post-3R review process will need to be established. It is expected to be quarterly, speedier and more flexible than the past process.

During discussions about the new Toolkit, and what training would be needed for implementation, the broader point was made that cataloguers need training to support cataloguer judgement. Cataloguers who have worked in a very prescriptive environment need training to have the confidence in their judgement when they no longer have the prescriptive environment.

A Hungarian FRBRisation project, http://opac3.frbr.monguz.hu/ aims to link titles for the purposes of copyright information, so that titles could be legally provided online. There are additional copyright holders at every level of the Work, Expression, and Manifestation. An interesting aspect of the project was the tool developed to enable cataloguers to edit the relationships between entities – a graphical interface, used when the automatic matching was inadequate.

### BIBFRAME

Leif Andresen from the Royal Danish Library presented on BIBFRAME (BF) in Europe. He argued that all modern discovery needs to be entity-based, and observed that while many catalogues have a user interface which is entity-based (FRBR-ised), there have been no such user interfaces for cataloguers. Leif sees BF as having a future which includes record exchange, but does not think that BF will be the answer for making library data communicate with the 'outside' world. European users of BF would like it to be better correlated with RDA. However, RDA is format neutral, so it's unclear how BF users will achieve stronger connection to RDA. The next European BF workshop will be in Stockholm, September 17-18 2019.

## RDA implementation at Cambridge University

**Charlotte Smith, Rayleigh Library, University of Cambridge**

In the summer of 2013, the University of Cambridge began implementing a programme of training across the 100+ libraries that make up the institution's library and information provision. Staff members from all of these various libraries made their way to a variety of venues over the course of several months to hear a group of trainers talk about the new international standard for descriptive cataloguing: Resource Description and Access. There was excitement. There was fear. There were lots and lots of questions.

Resource Description and Access, or RDA for short, was a fairly new concept in 2013. As one of the forerunners of RDA implementation, Cambridge was keen to train all library staff in its use. At that time, I was in post as a subject librarian at the Modern and Medieval Languages Library, managing the Germanic and Film collections. As bibliographic standards have always been a professional interest of mine, I was approached to form part of a group of library staff who would write and deliver the training materials. I was delighted to accept and we began working together at planning meetings shortly afterwards.

At this point, it might be useful to have some more background about the libraries system within the university. Essentially, there are three different types of library: the Main University Library, the Faculty and Department Libraries, and the College Libraries. Historically, the College Libraries collect resources and materials to support the first two years of undergraduate courses. They might also have postgraduate materials and special collections. Each Cambridge College has a library; these libraries exist to support the learning and teaching, and sometimes research, needs of College members. As such, people from other Colleges are not permitted to borrow resources. Each Faculty and Department also has a library, which provides learning resources, information skills teaching, and often research support for its members and the wider student community. The Main University Library is a copyright library; its core collections are research-based in nature and it sets the agenda in terms of policy and strategic development. It was considered essential that, in order to effect real change and encourage genuine buy-in from library staff across the different types of library, the training team be composed of people from across the three library categories. As such, I was representing a Faculty library, and I worked alongside several colleagues from College Libraries and the Main University Library. It was hoped that this collaborative non-hierarchical approach would avoid any perceptions of top-down authoritarian control in terms of the future of cataloguing standards within Cambridge. Indeed, as a Faculty representative, I was able to write one of the training modules myself. I was then able to gain feedback from my colleagues within the training group, all of which was useful, before presenting these materials to the wider community.

After several planning meetings where we discussed the size and scope of the training to be delivered, fed back on the modules several of us had written and analysed the comments from the pilot session, it was time to start the training for real. Over the course of several months, we ran and facilitated training sessions for over a hundred library staff members working in many different types of library. Even within the three categories of library in Cambridge mentioned above, there exists a huge wealth of differences. For example, in one session, participants ranged from a rare books librarian to a scientific information manager. Although RDA covers all descriptive cataloguing, the focus and the actual data that someone finds important in a bibliographic record varies considerably.

We decided to create and run two modules. Each participant was expected to attend both sessions, one for module one and one for module two. The first encompassed a brief overview of the need for RDA, the basic descriptive differences between RDA and the Anglo-American Cataloguing Rules (AACR2), and the actual changes in the MARC21 coding.

The second dealt with the slightly more complex and unique features of RDA, namely uniform titles, translations, and relationship designators. Many of the library staff members who attended these sessions had direct experience of AARC2 and MARC21 but little or no knowledge of other types of information standards. Therefore, concepts such as relationship designators were very new and their value was not fully explored due to time constraints.

We ran several sessions a week for three months. Although this took time away from our main roles, it was manageable because we spread the work evenly amongst the trainer team. It provided us all with a unique experience in terms of working in a non-hierarchical cross-sector team and it strengthened my teaching skills to the point where I felt extremely confident when training and presenting to groups. Overall, participants indicated in their feedback that they found the training useful and interesting. However, as with any new international standard, there was an amount of reluctance to embrace RDA as the new way of cataloguing primarily due to the perceived amount of work involved when creating bibliographic records. For example, no longer having the rule of three for co-authors requires cataloguers to create additional added author fields with relationship designators. Although helpful for our users, several attendees were anxious about the amount of staff time that changes like this might demand.

Fast forward six years to 2019 and RDA is now, from my point of view, business as usual within the university. I took a career break between 2015 and 2018 for personal reasons, but after re-joining the university as a science librarian in 2018, I can see that the standard that at first appeared time heavy and exotic to some has taken on the mantle of the new normal. I think what has helped is the fact that several other changes have also taken place over the last few years. Although at first our policy stated that only original cataloguing had to be carried out in RDA, many libraries decided to look primarily for RDA records when downloading from external sources. As time has passed and more institutions are using RDA, the amount of RDA records in circulation for new books is increasing while those in AACR2 diminish. What has also helped immensely is the introduction of our new university-wide library management system, ALMA. Before ALMA, the libraries were split into several databases, each having their own set of bibliographic records. With the advent of the new library management system, we now have all the libraries' bibliographic records in one place. In practical terms, this means that we now only have to create one bibliographic record for a resource within Cambridge. All that extra work that some feared would result from the changes that RDA makes to descriptive cataloguing is diminished by this new development in data management software, at least for modern collections. This I suspect has played in huge part in quelling opposition to RDA.

What are the lessons that we can learn? Firstly, I believe that it is crucial to have a cross-sector team working together to create and deliver this type of training in order to gain as much genuine buy-in from your community as possible. There will always be anxiety and worry in a time of change; being a trainer is a good opportunity to listen and to discuss these concerns. We had a huge amount of questions at our sessions, not all of which we could answer on the spot, but we always signposted to appropriate support. One of the main criticisms that we received was that each participant had to attend two sessions. We thought carefully about how we could have done this differently, but the volume of material we had to cover meant that it would not have been possible to only have one session. If we were delivering this training in 2019, I would suggest that the first module be delivered online, before moving on to present the second module in person with time for questions. Such a blended approach would perhaps reassure people's anxieties over time management. It might also have been a better idea to theme the training sessions by discipline rather than let everyone sign up to any of the sessions. We could have run sessions aimed at Arts and Humanities librarians, other for STEM librarians, etc. As a previous cataloguer of film and audio-visual materials, I always hoped that these formats would be more thoroughly investigated, although this was not possible due to our own time constraints as trainers.

To conclude, implementing a new standard for cataloguing is always going to be an interesting process. Time constraints, perceived lack of value, anxiety about change; these all lead to a challenging time for any trainer or implementer. What I would genuinely say is that being a trainer during the process of RDA implementation at Cambridge is one of the highlights of my professional career. It taught me so much in terms of professional resilience, confidence, being a teacher who listens as well as talks, and of course it deepened my knowledge of the principles and rationale of RDA itself. Returning to the university and its library community in 2018, and seeing those changes in how we catalogue resources becoming the new normal, fills me with tremendous pride.

# The Dublin Core Metadata Initiative (DCMI): A new phase

**Paul Walk, DCMI Managing Director**

## Introduction

The Dublin Core traces its origins to a workshop in Dublin, Ohio in 1995. The meeting, convened by OCLC and the National Center for Supercomputing Applications (NCSA), discussed how a core set of semantics for Web-based resources could be applied to Web resources to facilitate easier search and retrieval. The output from this meeting was named "Dublin Core metadata", based on the location of the workshop. With the rapid growth of the Web a large community quickly formed around Dublin Core, and this was formalised as the Dublin Core Metadata Initiative (DCMI).

DCMI has continued to run annual workshops - later to grow into full conferences - in an unbroken line back to 1995. The community has grown and diversified, and the Dublin Core standard has evolved - from the classic fifteen elements of the *Dublin Core Metadata Element Set* [1] to the expanded and more semantically useful *DCMI Metadata Terms*.[2] Dublin Core metadata is now very widely adopted, in a broad range of domains and use-cases.

Other recommendations and other supporting documents have been produced over the years, and the community has engaged in many metadata-related research and development initiatives, including significant contributions to the development of concepts such as the "metadata application profile".

For a period DCMI was incorporated as a Singapore-based legal entity; currently, however, DCMI is officially a project of The Association for Information Science and Technology (ASIS&T).[3]

Since mid 2017, DCMI has undergone a careful process of reviewing itself as an organisation, with the intention that it be re-aligned to address more recent challenges in metadata. This article describes the outcome of this process, and describes some recent changes and initiatives in DCMI, with a focus on those which might be of particular interest to the library community.

## DCMI renewed

DCMI is, and has always been, a membership organisation. For many years this included a mixture of organisational as well as individual paid-membership. However, the organisation has recently discontinued the individual membership in favour of developing a sustainable future focussed on addressing the agendas of its organisational members. To this end DCMI has a Governing Board[4] comprised of representatives from each of its members, which oversees the strategic direction of the organisation.

---

1. "DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description." Accessed May 27, 2019. http://www.dublincore.org/specifications/dublin-core/dces/.

2. "DCMI: DCMI Metadata Terms." Accessed May 27, 2019. http://www.dublincore.org/specifications/dublin-core/dcmi-terms/.

3. "ASIS&T." *ASIS&T* (blog). Accessed May 27, 2019. https://www.asist.org/.

4. "DCMI: DCMI Governing Board." Accessed May 27, 2019. http://www.dublincore.org/groups/governing-board/.

The renewed strategy for DCMI is directly expressed through its website, relaunched in early 2019.[5] DCMI has one of the older websites in the world - one which has accumulated an enormous amount of material such as specification documents, outputs from research and development, notes from meetings etc. The relaunched website is the product of an eighteen-month redevelopment project, where the resulting site is aligned to the current strategic priorities agreed by the Governing Board.

The activities of DCMI are now categorised under four high-level "themes": *stewardship*, *community*, *learning* and *development*. Each of these is briefly described here.

**Stewardship**

Possibly the most important of DCMI's activities can be grouped under the theme of stewardship. DCMI takes stewardship very seriously. In an increasingly networked world, vocabularies such as the DCMI Metadata Terms are relied upon as a digital *infrastructure*. The Linked Open Vocabularies website,[6] which collates the existence and use of controlled vocabularies in lined data positions the DCMI Metadata Terms as the single most used vocabulary (see figure 1).

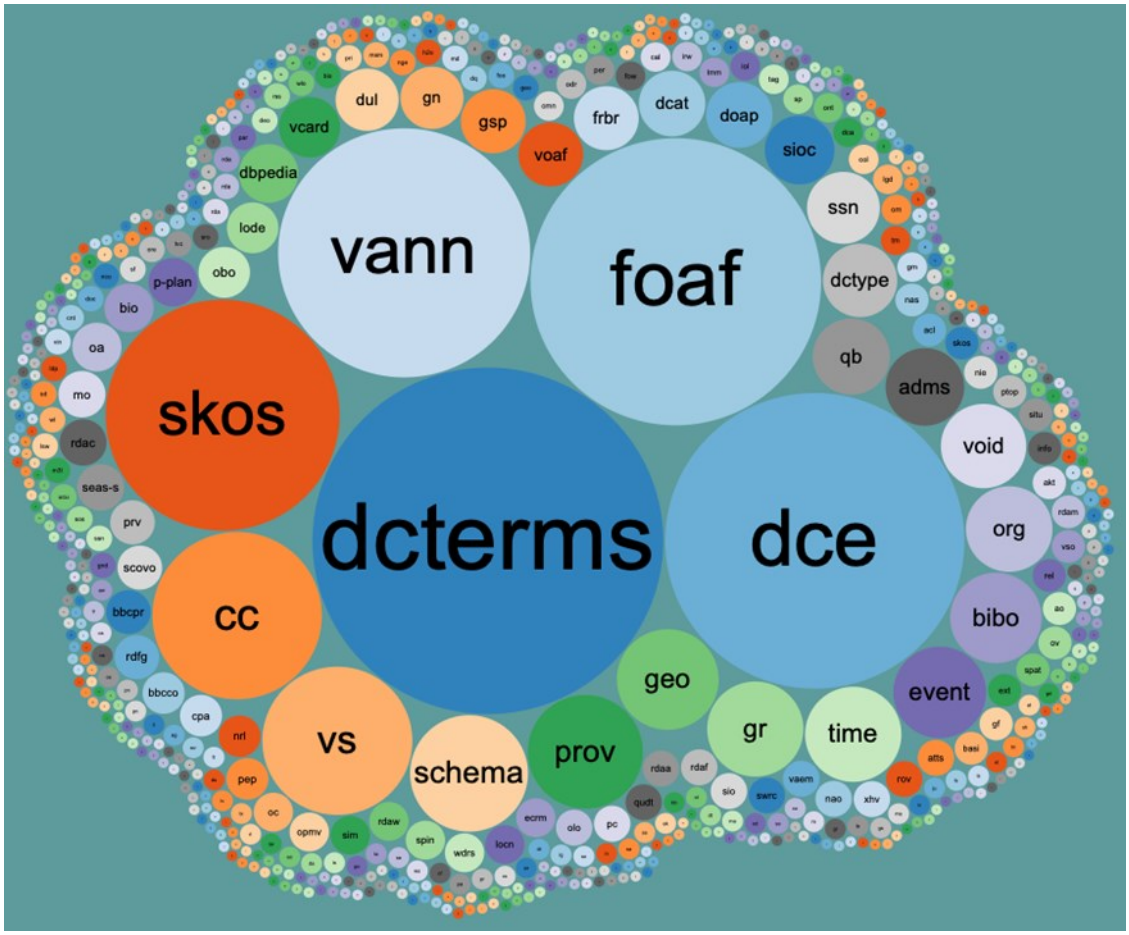


**Fig. 1: Linked Open Vocabularies**

5. "DCMI: Home." Accessed May 27, 2019. http://www.dublincore.org/.

6."DCMI: Agreement between DCMI and the FOAF Project." Accessed May 28, 2019. http://www.dublincore.org/collaborations/foaf/good*neighbour*agreement/.

Much of our digital infrastructure is maintained by communities of interest, relying on volunteer effort, with little guarantee of sustainability. DCMI works to mitigate this through its paid-membership model, relying on the support of those institutions which value the continued development and maintenance of the specifications under DCMI stewardship.

DCMI offers different levels of stewardship. The most basic is what we have come to call the "Good Neighbour Agreement". DCMI has such an arrangement with the *Friend of a Friend (FOAF)* specification.[7] Essentially, this consists of a mutual agreement to maintain each others' "namespaces" (i.e. the domain name registrations) in the event of either of the organisations failing, and to provide a minimal level of curation of the specification documents, ensuring that they continue to be available on the Web.

Other, more comprehensive levels of stewardship are also offered. For example, DCMI is responsible for continuity of access to the *Learning Resource Metadata Initiative (LRMI)* specification,[8] maintaining the namespaces and hosting all of the documentation as well as the machine-readable specification itself. This also entails maintaining the PURL redirects used to ensure persistence of the URLs. LRMI continues to be developed by an active community. DCMI does not interfere in this, but requires a representative from that community to be a member of the committee responsible for all DCMI's stewardship activity - known for historic reasons as the *Usage Board.*[9]

More recently, DCMI has assumed stewardship responsibility for the *Bibliographic Ontology (BIBO*).[10] In this case, there is no active development community so DCMI has taken complete responsibility for this, working with the authors of the ontology, one of whom now serves on the DCMI Usage Board as the representative for BIBO. If there is evidence of demand, then DCMI will convene a group to explore potential future development for BIBO. If there is no such demand, then DCMI can continue to ensure continuity of availability of the specification, documentation and namespace.

**Community**

DCMI is, at its heart, a large community of metadata researchers and practitioners. All of the significant outputs from DCMI's activities over more than twenty years have come from this community. In fact, DCMI has often helped to bridge between different communities of metadata experts. DCMI is also truly international, with members from Europe, North and Latin America and Asia. Some of the more important DCMI specifications are translated into other languages - notable Spanish, Portuguese and Chinese.

The sheer range of domains and interests represented in the DCMI community can be seen in the long list of historic community groups.[11] As part of the new strategy, DCMI has adopted a model of overarching *interest groups,*[12] aimed at establishing communities of interest, which may spawn more focused, time-limited *working groups*, tasked with specific goals. Some recently created examples include the *Application Profiles IG*, *Automated Subject Indexing IG*, and the *Teaching Metadata IG*.

---

7. "DCMI: Agreement between DCMI and the FOAF Project." Accessed May 28, 2019. http://www.dublincore.org/collaborations/foaf/good*neighbour*agreement/.

8. "DCMI: LRMI." Accessed May 28, 2019. http://www.dublincore.org/specifications/lrmi/.

9. "DCMI: DCMI Usage Board." Accessed May 28, 2019. http://www.dublincore.org/groups/usage-board/.

10. "DCMI: BIBO." Accessed May 28, 2019. http://www.dublincore.org/specifications/bibo/

11."DCMI: DCMI Community Groups." Accessed May 28, 2019. http://www.dublincore.org/groups/.

12. "DCMI: Community." Accessed May 28, 2019. http://www.dublincore.org/themes/community/

As well as the specific groups, the DCMI community expresses itself through two other notable activities.

Firstly, for a number of years, DCMI has run webinars,[13] written and presented by members of the community. The webinars are nearly always focussed on some aspect of good metadata practice, with some (but certainly not all) being quite technical. These webinars have proven to be very popular in the past. DCMI is now planning a more sustained programme of webinars, working with ASIS&T to provide the infrastructure. In keeping with our international nature, several recent webinars have been conducted in other languages - sometimes repeated in, for example, both English and Spanish.

Secondly, DCMI has continued to arrange an annual meeting - now a conference - every year since the original workshop in 1995.[14] The conferences are planned in a rough cycle moving between Europe, the Americas and Asia, in order to be accessible to as much of the global community as possible. In September 2019, the conference will be hosted by the National Library of Korea in Seoul.[15]

**Learning**

Teaching and learning metadata-related theory and practice is a key concern of a significant part of the wider DCMI community. DCMI has developed and published resources to support teaching metadata, and there is evidence to show that these continue to be widely used and referenced in the curricula of information management courses around the world.

Recognising the continued demand for such resources, DCMI has recently updated several of them, notably the widely cited *Metadata Basics* web page[16] and the *Dublin Core User Guide.*[17]

A more recent addition to the set of resources supporting teaching and learning is the *Linked Data Competency Index (LDCI*).[18]This is a categorised set of statements about the knowledge and skills required to work with linked data. The LDCI provides a "map" of linked data competencies aimed at both independent learners wishing to learn linked data methods and technology, and for teachers who want to design courses on the subject. DCMI is currently examine ways on which it might use this as a framework to contextualise other relevant learning resources.

The interest in teaching metadata seems to be undiminished in the DCMI community - in fact, there is some evidence to suggest it has grown. Because of this, DCMI has recently convened a new interest group - the *Teaching Metadata IG*[19] which will explore this as a general area off interest, as well as taking care of some of DCMI's learning resources and the Webinar programme.

---

13. "DCMI: Webinars." Accessed May 28, 2019. http://www.dublincore.org/webinars/.

14. "DCMI: DCMI Annual Conferences." Accessed May 28, 2019. http://www.dublincore.org/conferences/.

15. "DCMI: DCMI 2019 Seoul." Accessed May 28, 2019. http://www.dublincore.org/conferences/2019/.

16. "DCMI: Metadata Basics." Accessed May 29, 2019. http://www.dublincore.org/resources/metadata-basics/.

17. "DCMI: Dublin Core User Guide." Accessed May 29, 2019. http://www.dublincore.org/resources/userguide/.

18. "About - Linked Data Competency Index." Accessed May 29, 2019. https://dcmi.github.io/ldci/.

19. "DCMI: Teaching Metadata IG." Accessed May 29, 2019. http://www.dublincore.org/groups/teaching*metadata*ig/.

**Development**

In addition to its support activities, DCMI has often provided the "venue" for technical development and innovation activities, sometimes in collaboration with other communities. Following the development of the ubiquitous DCMI Metadata Terms, the community has in more recent years focused on the concept of the metadata application profile, developing supporting frameworks and conceptual models such as the Singapore Framework.

Very recently, DCMI has returned to this space, with the establishment of a new *Application Profiles Interest Group*.[20] Despite only having started in April 2019, this group has already been very active, having assembled an impressive set of use-cases and requirements in order to focus the discussion. Ultimately, this group intends to develop a "core vocabulary of application profile components and constraints".

**Get Involved!**

There are many ways in which organisations and individuals can participate in the DCMI community and its activities. The various community groups are a good starting point - they all have mailing lists which allow anyone to join. While active contribution is encouraged of course, it is also perfectly acceptable to 'lurk', absorbing the discussions which happen there. In some cases, such as the Application Profiles IG where the community has already found a shared endeavour requiring more structured discussion, other channels - notably Github "issues" - are used instead of mailing lists. In any case, these channels are open to all - DCMI is, and always has been, a very open community.

DCMI is always looking for interesting topics for webinars - especially right now as we develop a new webinar programme.

The annual conference is always a great opportunity to engage with the DCMI community - and this year's conference in Seoul promises to be exciting with some new initiatives - such as a Wikidata tutorial and "hack day".[21]

In April 2019, DCMI was very pleased to welcome its newest institutional member - *Library and Archives Canada (LAC).*[22] Around half of DCMI's thirteen institutional members[23] are libraries (including three national libraries) or library-focused organisations, demonstrating that the library sector recognises the value of supporting and steering DCMI (all DCMI institutional members gain a place at the table of the DCMI Governing Board). DCMI invites all organisations with an interest in supporting the development and use of metadata to consider joining as members. With a healthy level of membership DCMI, is more confidently able to provide continuity of access to and maintenance of some globally-important metadata specifications, to continue to develop learning resources to aid in their use, and to foster the development and innovation which will take metadata forward.

Please address enquiries to the DCMI Executive at: contact@dublincore.net

---

20. "DCMI: Application Profiles IG." Accessed May 29, 2019. http://www.dublincore.org/groups/application*profiles*ig/.

21. "DCMI: DC-2019 Hack Day - Call for Participation." Accessed May 29, 2019. http://www.dublincore.org/conferences/2019/hackday/.

22. Canada, Library and Archives. "Library and Archives Canada (LAC)," September 11, 2018. http://www.bac-lac.gc.ca/eng/Pages/home.aspx.

23. "DCMI: Organisational Members." Accessed May 29, 2019. http://www.dublincore.org/members/.

**EDItEUR**

**Graham Bell, Executive Director and Chief Data Architect, EDItEUR**

Cataloguing in a library, shelving in a bookshop? Different, but with so much in common.

Both rely critically on metadata. In the case of a library, ensuring patrons can find the resources they need is all about the catalogue, full of in-depth, authoritative metadata about the physical and digital holdings. In a store, physical shelving is often driven by subject, by publication date (those 'new books' shelves), and so on. Online, a bookstore depends on metadata to drive discoverability and purchase. And whether it is in a library, or in a store, standards underpin the exchange of metadata across the information and publishing sector. But although on the surface, the data needs might be thought to be very similar, the data used in libraries and data bookstores looks entirely different, the supply chains – both for the books and the data – are quite different, and the demands on store and library staff are different too.

Focusing on the data itself illustrates some of the contrasts. Library metadata standards like MARC or RDA are well-established, but aren't used widely in the publishing world, other than by a few academic publishers or suppliers whose main customers are libraries. Publishers in North America, across Europe and increasingly in the Asia-Pacific region predominantly use a standard called ‘ONIX’ to distribute data about their book products to bookstores. Libraries use Dewey or UDC for subject classification, whereas publishers and booksellers use a range of subject schemes – ‘BIC’, ‘BISAC’ or a host of other country-specific schemes, and increasingly a global scheme called ‘*Thema*’.

So who in the commercial publishing and bookselling world creates ONIX or BISAC, CLIL or *Thema*? These are ‘trade standards’, the prosaic ‘plumbing’ of the data supply chain on which the industry depends, defined and promoted by trade organisations for the benefit of the industry as a whole.

EDItEUR (https://www.editeur.org/) is one such organisation. London-based and with just three and a half staff, it is – like most similar organisations – a not-for-profit, and is best known today for creating, supporting and promoting ONIX for Books, an XML-based message format used widely for communicating metadata about books and book-like products between publishers and booksellers.

Why is ONIX different from MARC, when both broadly describe the same books? Part of the answer lies in the contrast between the librarian’s ‘resource’ and the publisher’s ‘product’. The latter has an overtly commercial outlook, and part of the aim is to maximise sales of those products. ONIX as a standard recognises this – so rather than remaining a purely bibliographic description of the book and its content, ONIX also encompasses a range of marketing, rights and commercial information that would be of little importance to the librarian, but which are vital in reducing ‘friction’ in the book, e-book and audio supply chains. Another contrast is seen in the dynamic nature of the commercial metadata. Publishers begin to distribute information about a new book several months in advance of publication, and continue to update, supplement and even revise that data, not only in the months leading up to the publication date but throughout the lifecycle of the book.

So ONIX has a range of commercial stakeholders, from publishing companies of all sizes to book retailers, data aggregators and logistics organisations. EDItEUR’s membership is built from these stakeholders. Their commitment is a sign that they support an open, standards-based data supply chain. Through direct contact with EDItEUR, members have a voice in the future direction of standards development. EDItEUR relies almost entirely on membership revenue. And yet EDItEUR makes the standards specifications available free of charge to all – because the best outcome for members is that *everyone else uses the same standards too*.

The development and governance process for ONIX ensures benefits flow broadly and without favouring particular commercial organisations: EDItEUR encourages its members around the globe to form 'national groups' representing the whole data supply chain in that country, and ultimately, it is these national groups rather than EDItEUR that have the final say on developments to the standard.

That development and governance model, emphasising wide participation in national user groups, EDItEUR membership to pay the bills, and free availability of the standard for all, is also used for *Thema*, EDItEUR's global subject category scheme. Widely used library-focused schemes such as Dewey or UDC are both too complex and too limiting for commercial use, and the book business has developed numerous nationally-focussed subject schemes. BISAC is one such, developed and promoted by the New York-based Book Industry Study Group, and *Thèmes CLIL* is its quite different French equivalent, but there are many tens of other national schemes creating a babel of competing schemes in the international book trade. *Thema* is an attempt to create a commercially-relevant subject scheme suitable for global use, and it has grown fast: it's just five years old, but has already supplanted established national schemes in many European countries. In the UK, *Thema* is best viewed as 'the latest update of the previous BIC subject scheme', and as a development from the BIC scheme is easy to adopt for UK publishers and retailers.

For EDItEUR, the international dimension to standards is critical: our industry is increasingly international in scope, our members are spread internationally, and publishing and bookselling issues are often common across many countries. The standards themselves reflect this:

```
<Contributor>
   <SequenceNumber>1</SequenceNumber>
   <ContributorRole>A01</ContributorRole>  <!-- author -->
   <NameIdentifier>
      <NameIDType>16</NameIDType>         <!-- ISNI -->
      <IDValue>0000000083888172</IDValue>
   </NameIdentifier>
   <PersonName>서영은</PersonName>
   <KeyNames>서</KeyNames>                 <!-- Seo -->
   <NamesAfterKey>영은</NamesAfterKey>       <!-- Yeong-eun -->
   <BiographicalNote textformat="05"><p>서영은 강원도 강릉 출생이며, 강원도 강릉사범학교를 졸업하
였으며 건국대학교 영어영문학과를 중퇴하였다.</p><p>자아의 갈등을 상징적이고 우화적인 수법으로
그려내는 리얼리즘 계열의 작가로 알려져 있으며, 현실에서의 삶의 조건에 대한 보다 인간적인 관찰을 거
쳐 초월적 입장에서 어떤 절대세계에 의미를 부여하고자 했다.</p></BiographicalNote>
</Contributor>
```

The example extract of metadata describes a Korean writer, Seo Yeong-eun (**서영은**) – and you'll note that the XML-based format of ONIX works as well in Hangul, or Arabic or Cyrillic, as it does with Latin. It can include text in many languages, in parallel if necessary. The standard specification has been adopted as a national standard in China (GB/T 30330), and is also the preferred format for delivery from publishers to Amazon, Flipkart (in India), BOL (in The Netherlands), or to Waterstone's online stores. But the stakeholders are not all huge organisations – EDItEUR's smallest members are two-person companies, who see the standard as levelling the playing field.

Equally, *Thema* is intended for global use. The whole multilingual scheme can be viewed at https://ns.editeur.org/thema. It's not intended as a replacement for Dewey or UDC and lacks the detail and rigour demanded by academic libraries – but it might be a good fit for the public library service. A book like Colson Whitehead's *The Underground Railroad* might have the *Thema* categorisation:

FV – Historical fiction
FBA – Modern and contemporary fiction
NHTS – Slavery and abolition of slavery
1KBB-US-SE – US South Atlantic states
3MNB – 1820s
5PB-US-C – Relating to African American people

This provides multiple ways of accessing the book, whether via a facetted search or a more traditional 'browse by subject'. The codes (*eg* 'FBA') are independent of language, whereas the literals ('Modern and contemporary fiction') are available in around 20 languages so far with others on the way. MKE means the subject is dentistry, but also *zahnheilkunde*, *fogászat*, طب الأسنانand *牙科*.

So if these standards have an overtly commercial tenor, where do they intersect with library practices? ONIX metadata records are usually distributed widely across the book trade several months prior to publication, so that ONIX data is often used to create an initial skeleton MARC record. The British Library's CIP programme (operated by BDS Ltd.) works in this way, and the US Library of Congress is actively soliciting more ONIX data from publishers. The MARC data may later be enhanced (or even *corrected*, for publishers are less bound by cataloguing rules than are librarians) via a book-in-hand process, but its genesis frequently lies in ONIX. ONIX of course uses the same range of identifiers and authorities as library catalogues, so interoperability is good. Each product described has an ISBN, contributors can be listed with their ISNI or ORCID identifiers, and there is a publicly-available mapping procedure developed by OCLC to convert from ONIX to MARC (MARC to ONIX is less successful, because the commercial data that provides the value in ONIX would be missing).

In addition to developing and supporting its own standards like ONIX and *Thema*, EDItEUR also provides management support to *other* identifier standard organisations, in particular the ISNI International Agency (www.isni.org) and the International DOI Foundation (https://www.doi.org/) which manage the ISNI and DOI systems on behalf of the ISO. Until recently it provided similar support to the International ISBN Agency. All this, combined with EDItEUR's global outlook and spread of members and contacts, gives EDItEUR an unparalleled insight into the use of identifier and metadata standards across the industry, and this expertise can be called upon by EDItEUR members when necessary.

EDItEUR in summary

- a global trade standards body for the book, e-book and audiobook sector
- not for profit, membership-supported
- responsible for ONIX, *Thema,* EDItX and EDIFACT standards
- provides management services for ISNI, DOI agencies

For more details see https://www.editeur.org or e-mail info@editeur.org

Graham Bell, Executive Director, EDItEUR (ISNI: 0000000427566266)

Graham is Executive Director of EDItEUR, responsible for the overall development of EDItEUR's standards and the management services it provides on behalf of other standards agencies. He joined EDItEUR as its Chief Data Architect in 2010, focussed on the continuing development and application of ONIX for Books, and on other EDItEUR standards for both the book and serials sectors. Graham previously worked in IT for HarperCollins Publishers in the UK, and as an Editor in the magazine industry with Redwood Publishing and BBC Magazines.

## Foundations for the future: The British Library's Collection Metadata Strategy, 2019-2023

**Alan Danskin, British Library**

*Foundations for the future*, the British Library's Collection Metadata Strategy was published in April 2019. *Foundations for the future* sets out the Library's metadata strategy to 2023.

*"Collection metadata* is an umbrella term for the structured information required to efficiently manage, access and preserve the collection. It is a key organisational asset, representing centuries of resource investment. Efficient exploitation and stewardship of collection metadata underpins delivery of The British Library's core purposes, strategies and portfolios."[1]

"Our vision is that by 2023 the Library's collection metadata assets will be unified on a single, sustainable, standards-based infrastructure offering improved options for access, collaboration and open reuse"

Strategic activities will be guided by the following priorities:

- Enhance the Library's ability to exploit its collection metadata assets in order to deliver its strategic priorities, portfolios and programmes
- Ensure that the value of the Library's Collection Metadata investment is maintained through effective stewardship
- Enable open access to collection metadata in order to improve resource discovery and promote wider community reuse

The period covered will see the Library facing several key metadata management challenges.

- Unifying Collection Metadata Infrastructure
- Maintaining and Developing the Value of Collection Metadata Assets
- Delivering Efficient and Sustainable Collection Metadata Processes
- Collection Metadata for Preservation, Rights and Management Information
- Spreading and Maintaining Best Practice

*Foundations for the future* combines continuity of purpose with our previous strategy, *Unlocking the value,*[2] with a shift of focus onto unification of the Library's metadata infrastructure.

Unification of the metadata infrastructure comprehends metadata standards; metadata creation systems, discovery and access. The opportunity arises from the need to replace the Digital Library System (DLS); the Integrated Archives and Manuscripts System (IAMS); the Sound and Moving Image System (SAMI) and the Aleph ILS over the next 4-5 years. New systems will be phased in by the Library Systems Transformation programme (LST).

---

1. British Library. *Foundations for the future: the British Library's collection metadata strategy, 2019-2023* https://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2019-2023.pdf

2. British Library. *Unlocking the value: the British Library's collection metadata strategy, 2015-2018* https://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2015-2018.pdf

By maintaining and developing the value of our metadata we ensure it becomes (and remains) fit for contemporary purposes.

The volume and complexity of demands for metadata exceed the capacity of traditional cataloguing processes. We will continue to remove barriers to efficiency and to develop workflows that combine machine processing with manual exception handling.

Traditional assumptions about metadata requirements for preservation and access are not valid for digital content and we have to respond to much more complex requirements.

The success of the Strategy will rely on effective communication to stakeholders and adoption and maintenance of best practices by Library staff.

**Strategic Priorities 2019-23**

**Enhance the Library's ability to exploit its collection metadata assets in order to deliver its strategic priorities, portfolios and programmes by:**

- Managing migration to a new unified infrastructure for collection metadata management to reduce operational costs and complexity
- Developing and maintaining innovative collection metadata creation and enhancement processes to support efficient exploitation of third party data sources e.g. automated entity extraction from full text, crowdsourcing and bulk record enhancement.
- Designing and supporting accurate and flexible rights management metadata for both content and descriptions to enable the Library to take advantage of new licensing options and increase reuse of the collections
- Providing colleagues with self-service tools, expert assistance and best practice guidance to improve consistency and efficiency in the exploitation of collection metadata and integrate with established training initiatives
- Ensuring requirements for new management information tools utilising collection metadata are included in LST procurement specifications to improve accurate assessment of collection strengths and weaknesses

**Ensure that the value of the Library's Collection Metadata investment is maintained through effective stewardship by**

- Collaborating with international, cross sectoral partners to ensure appropriate collection metadata standards are developed and implemented to support operational efficiency and preserve long-term value, e.g. creating opportunities to consume high-quality metadata from external sources
- Ensuring collection metadata accurately and persistently incorporates all necessary information to deliver the Library's portfolios and wider strategic objectives
- Developing processes to prioritise collection metadata enhancements that support delivery of improved services to internal and external users
- Managing long-term risk to the integrity of collection metadata against short-term efficiencies
- Undertaking collection metadata maintenance, enhancement and QA processes to support increased operational efficiencies and smooth transition to new systems
- Ensuring licensing of third party collection metadata is negotiated with a long-term view

**Enable open access to collection metadata in order to improve resource discovery and promote wider community reuse by:**

- Ensuring suitable collection metadata is available to enable and support new shared service initiatives, open research and commercial collaborations
- Increasing opportunities for interaction with our collection metadata to show the richness of the collection and its relevance to all
- Ensuring our collection metadata is accurately represented and accessible via all relevant national or global discovery channels and platforms
- Exploring opportunities for community engagement in metadata enhancement, e.g. via development and promotion of crowdsourcing challenges
- Taking steps to expose any collection metadata assets currently unavailable for resource discovery, partnership or collection management
- Ensuring comprehensive, accurate and timely rights metadata is available to support the sharing or purchase of any digital content
- Making sure our collection metadata is openly available in full compliance with current public sector best practice and related legislation

**Annual Implementation Plan**

Delivery of the strategy is tracked through a series of annual plans.

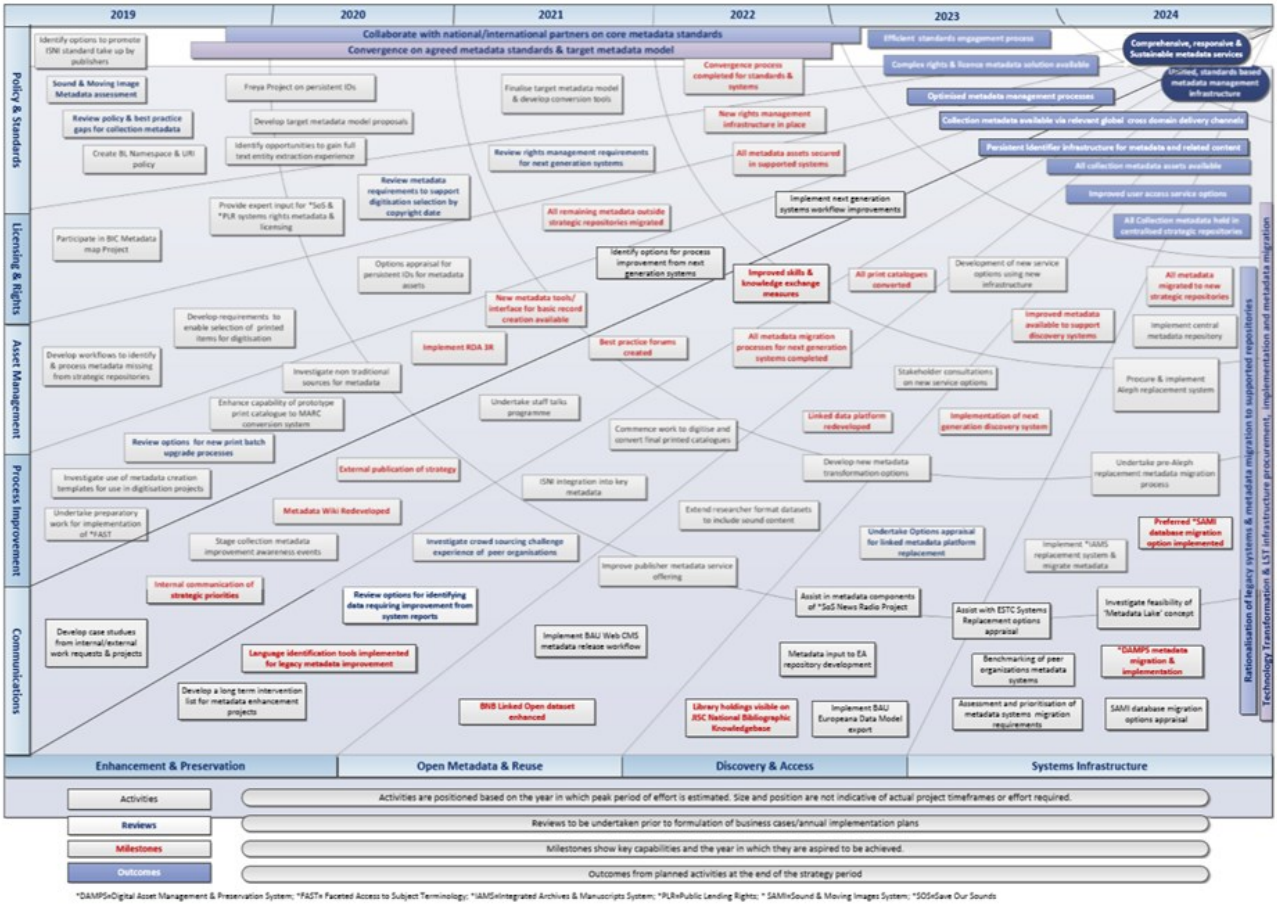


**Fig. 1 Road map Collection Metadata Strategy 2019-23**

Priorities for 2019-20 include:

- Metadata assessment, migration and configuration activities for the Digital Asset Management and Preservation System (DAMPS) and other LST Project strands
- A single target metadata model for use on next generation systems will be developed in consultation with stakeholders.
- Activities supporting publisher circulation of good quality metadata via trade supply chains will be investigated and piloted (e.g. use of the ISNI)
- Collection Metadata for Library holdings will be available via the new JISC National Bibliographic Knowledgebase

A graphical representation of the implementation plan has been published on the Library's website.[3]

**Unifying Collection Metadata Infrastructure**

The collection metadata infrastructure includes systems, metadata and business processes.

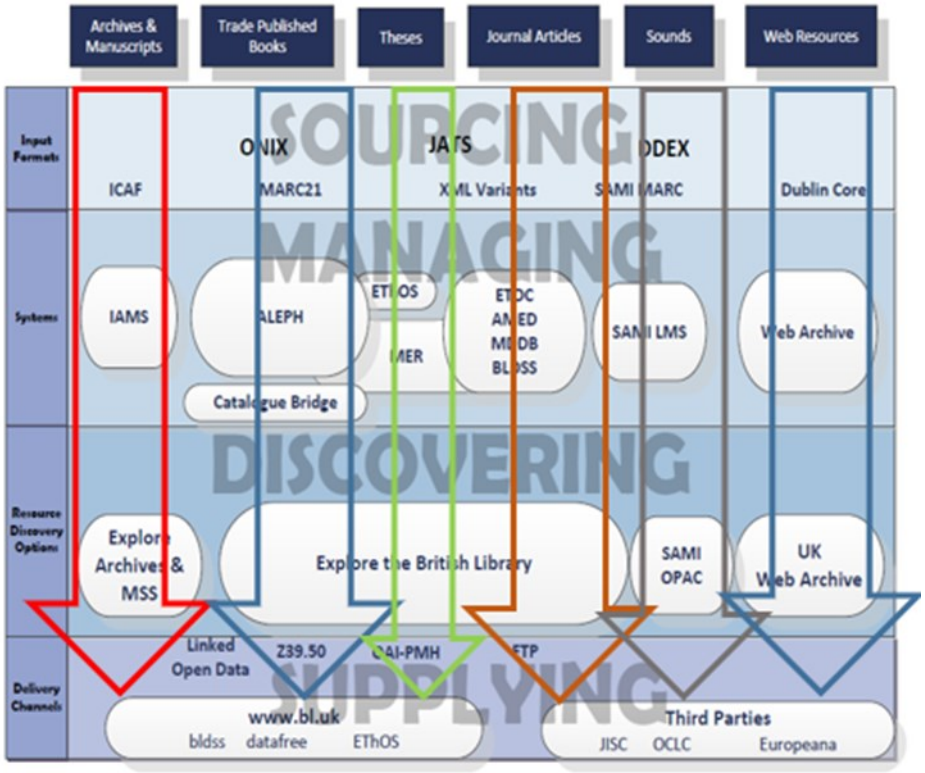


**Fig. 2 the metadata silos**

Separate workflows have developed into silos that cut across generic functions. The resulting infrastructure is increasingly difficult to maintain. The complexity and number of metadata transformations required are barriers to efficiency. These silos extend into the discovery layer, providing a fragmented view of the Library's collection.

3. Collection Metadata Strategy Roadmap 2019-2023 http://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-roadmap-2019-2023.pdf

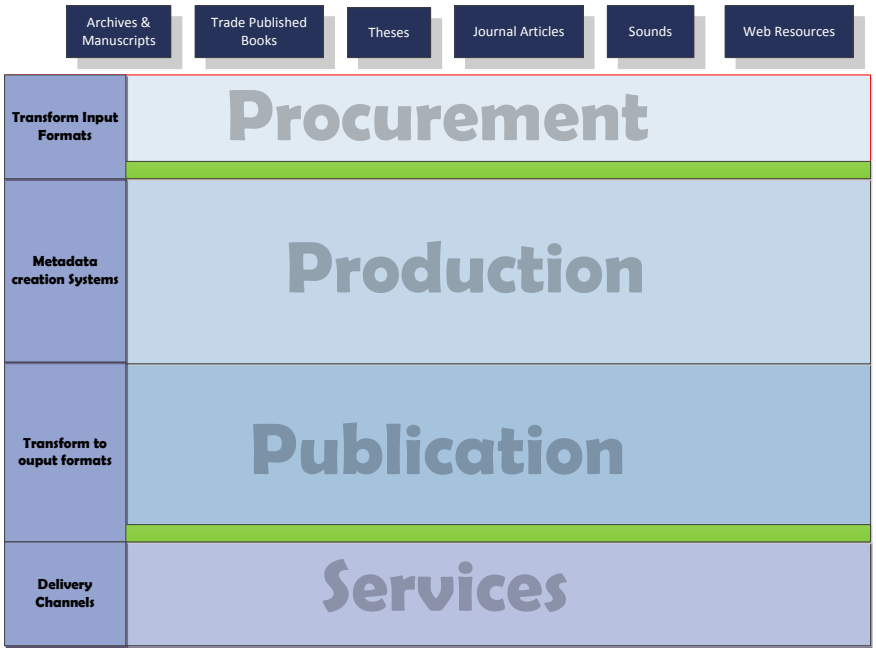


**Fig. 3 Unifying the infrastructure**

Unification will deliver a flexible infrastructure in which metadata is transformed on import to the library and on publication to services (the green lines indicate where transitions are required). No systems are illustrated on the diagram because the number and scope of replacement systems required is dependent on the outcome of future procurement exercises. The Library is currently working with Libnova to implement the Digital Assets Management and Preservation System (DAMPS). A proof of concept to migrate SAMI data to Aleph is planned for autumn, 2019. Requirements gathering to initiate the replacement of IAMS is also scheduled for 2019. It is anticipated that Aleph will be the last of the strategic metadata repositories to be replaced.

**Collection Metadata Standards**

Unification is expected to simplify and reduce the Library's standards portfolio. It has been explicitly recognised in the strategy that there is a dependency in some cases on system replacement. For example, we can't stop using MARC 21 until we replace Aleph.

But what *should* we do about MARC? In addition to MARC 21, we also use an internal variant, SAMIMARC, for our sound and vision resources. Our options will be constrained by the systems available in the marketplace, but the LST is our opportunity to replace MARC as an input format. We will continue to supply MARC data, even if we no longer use it internally. But, how will we replace MARC 21 and SAMIMARC? BIBFRAME is a linked data alternative to MARC, but the BIBFRAME model does not implement the IFLA Library Reference Model (LRM),[4] nor does it implement RDA.[5] In general, it would make more sense to implement a more expressive schema for import and export of less expressive schemata.

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

5. RDA: Resource Description and Access https://www.rdatoolkit.org

There are instances of competing standards. For example FAST (Faceted Application of Subject Terminology)[6] uses the same vocabulary as LCSH, but with a much simpler syntax. FAST is already in use for some workflows because it is easier and cheaper to implement. The Library has consulted[7] with the community on the feasibility of FAST as a replacement for LCSH. As a consequence of that consultation we have collaborated with OCLC and others to form the FAST Policy and Outreach Committee (FPOC).[8] FPOC will promote FAST and work with OCLC to develop and deliver an efficient service.

In addition to MARC and LCSH, we are reviewing standards in use for cataloguing unpublished resources, including archives and manuscripts. In general, we follow ISAD(G )[9] but specialised workflows are using either TEI[10]or an internal format to catalogue manuscripts. Maintaining three different input methods may provide for richer descriptions but represents a significant overhead. The solution once again, may be to provide a richer input format for all unpublished resources and transform data to Encoded Archival Description (EAD)[11] or TEI, as required, on export.

The Library remains committed to RDA. We are contributing to the implementation of the 3R[12] project, through EURIG,[13] and work is underway to develop supporting documentation and training materials. RDA 3R provides flexible options for how data elements and their provenance are recorded; concepts such as manifestation statement, aggregations and representative expression will prove valuable for integrating legacy data and metadata from other cultural heritage traditions. As an implementation of the Library Reference Model, 3R is aligned with the museum sector's Conceptual Reference Model (CRM)[14] and RDF (Resource Description Framework).[15] The LRM provides a rich set of entities many of which, such as Agent, Place, Timespan, can be reused independently of RDA.

**Target Metadata Model (TMM)**

The TMM is a framework model that will help us focus on the commonalities between our different metadata standards and models rather than obsess about their differences.

---

6.FAST https://www.oclc.org/en/fast.html

7. British Library. *Consultation on subject indexing and classification standards applied by the British Library*. https://www.bl.uk/bibliographic/pdfs/british-library-consultation-fast-abridged-dewey.pdf

8. FPOC https://www.oclc.org/en/fast/committee.html

9.International Council on Archives. G*eneral international archival description https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition*

10. Text Encoding Initiative (TEI) https://tei-c.org/

11. Encoded Archival Description https://www.loc.gov/ead

12. RDA FAQ 3R Project http://www.rda-rsc.org/node/551

13. European RDA Interest Group (EURIG) http://www.rda-rsc.org/europe

14. CIDOC CRM http://www.cidoc-crm.org/

15. W3C Resource Description Framework https://www.w3.org/2001/sw/wiki/RDF

The TMM developed from comparison of the Library's legacy metadata models with the IFLA Library Reference Model (LRM). The analysis identified 3 levels of metadata common to the different models.

- Level 1: metadata that puts an orderable unit into context (e.g. authority data)
- Level 2: metadata that describes an orderable unit (e.g. bibliographic description)
- Level 3: metadata that describes a dependent unit (e.g. analytic)

The TMM has to support comprehensive, coherent discovery of the collection; enable dissemination of collection metadata in the schemata required by stakeholders; and deliver efficiencies, e.g. by reducing the number of systems, workflows, standards and transformation to be maintained.

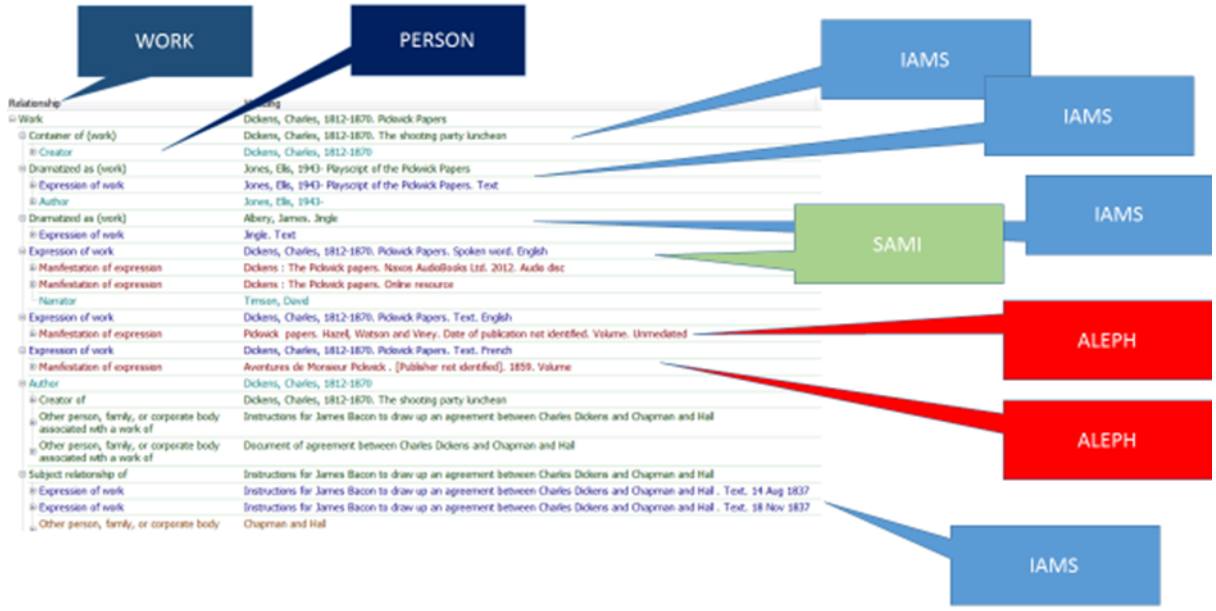


**Fig. 4 Converged results set for Pickwick papers in RIMMF 3**

Figure 4 illustrates a RIMMF3[16] "r-tree"[17] for Pickwick Papers by Charles Dickens. The "r tree" presents metadata for sound recordings, published books and unpublished papers taken from different silos in a unified RDA structure. Experimentation using RIMMF illustrated that while RDA offers potential for convergence around Levels 2 & 3, RDA is designed to contextualise them in relation to their content. This is not surprising but it is not clear that we could support a collection based view, as required by ISAD(G).

Initial work on the Target Metadata Model has mainly focused on reconciliation of RDA and ISAD(G). The TMM should be expressive enough to enable collection items to be shown in either context. The TMM won't be RDA or ISAD(G) or TEI or SAMIMARC, although cataloguers in specific workflow will be working with some of those standards.

16. RIMMF3 (RDA in Many Metadata Formats) http://marcofquality.com/wiki/rimmf3/doku.php?id=rimmf

17. An r-tree is a tool for visualising relationships between RDA entities

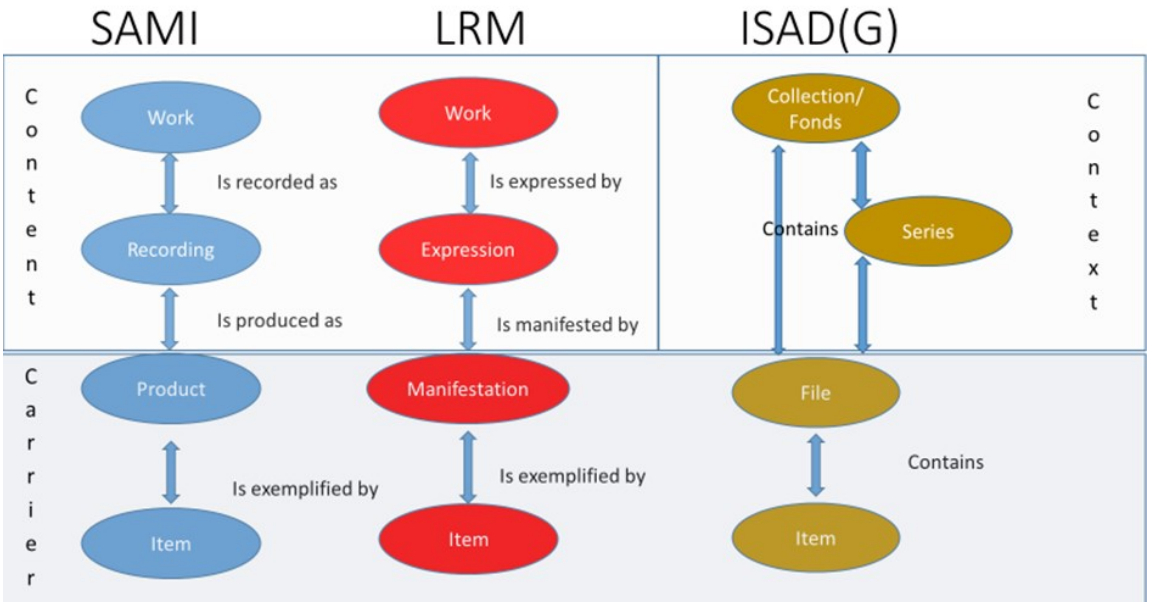


**Fig. 5 Matching models and entities**

Figure 5 Illustrates core entities and relationships in the SAMI model; the Library Reference Model, and ISAD (G). Although entities and attributes from different models are known by different names and may be recorded in different ways, they generally identify the same things, e.g. Persons, Items, Works, and Places.

Both LRM and SAMI models make a distinction between content and carrier and there is a relatively straightforward correlation between them. ISAD(G) makes provision for describing the carrier, but it is concerned with the context of items (if described) in archives or collections rather than their intellectual content.

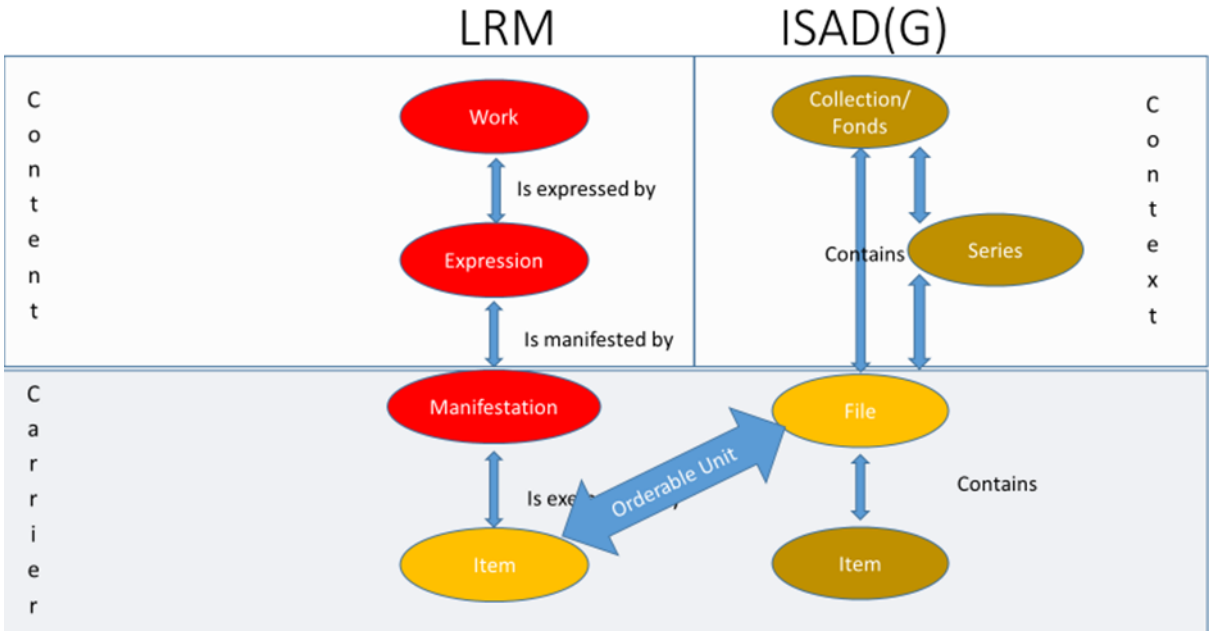


**Fig. 6 Convergence of models**

Figure 6 The SAMI data model is very similar to FRBR and can be merged into LRM. The Sound and Moving Image Collection also contains archival textual and graphic materials which may be more efficiently handled using archival standards.

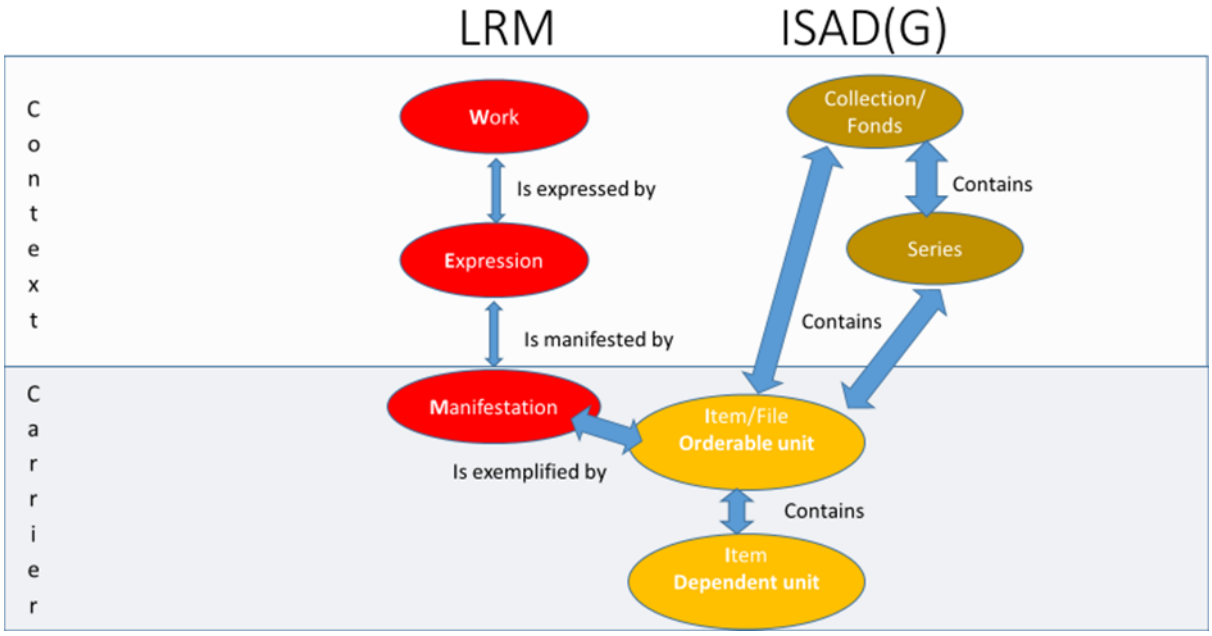


**Fig. 7 One collection 2 view points**

The LRM and ISAD(G) models offer alternative contexts for collection items. An LRM view brings together collection items with the same or related content.

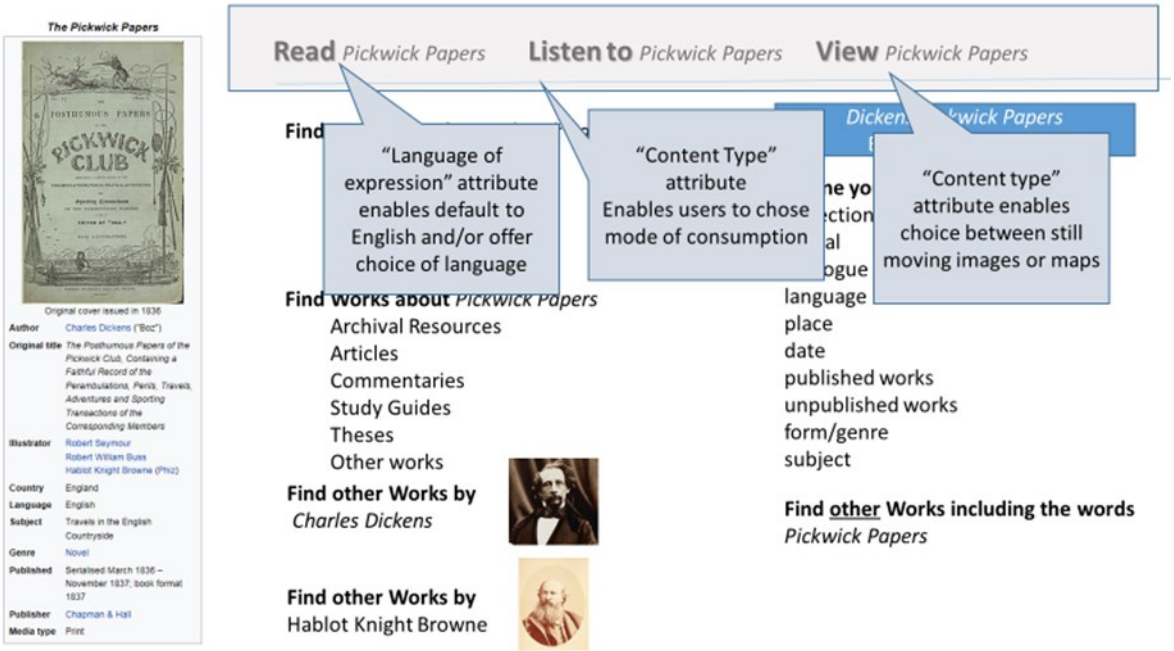


**Fig. 8 LRM view of Pickwick Papers**

An ISAD(G) view shows the place of a collection item in relation to the rest of the collection and it exposes collections that have not been fully analysed.

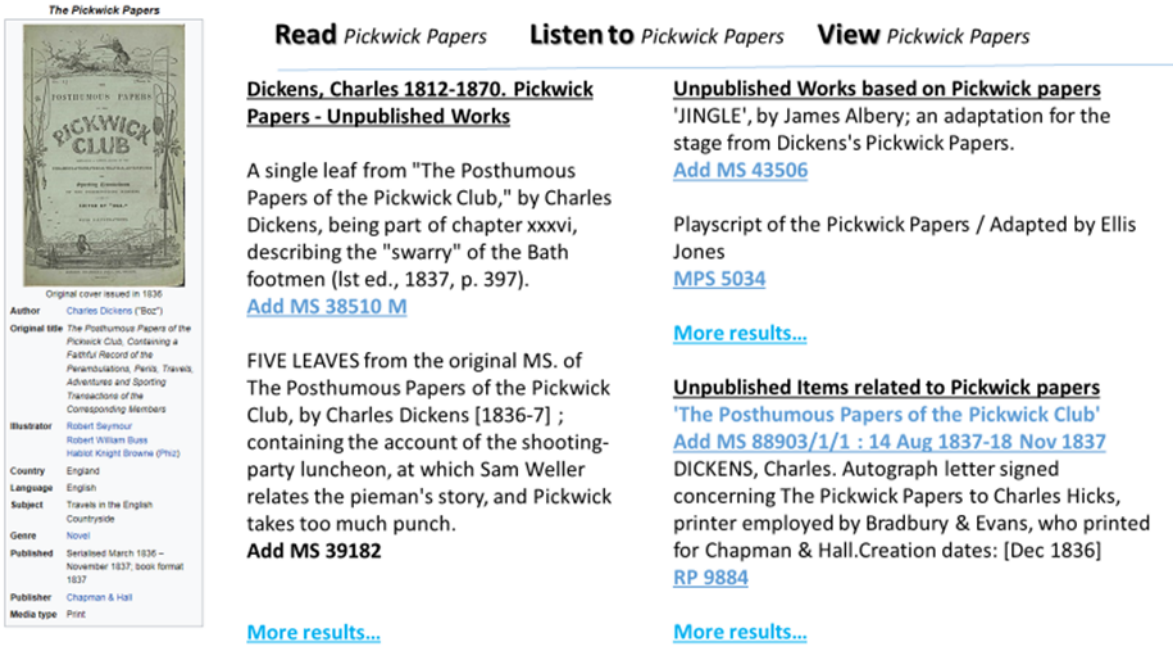


**Fig. 9 ISAD(G) view of unpublished resources related to Pickwick Papers**

## Conclusion

*Foundations for the future* continues the Library's collection metadata strategy, begun in 2015. In conjunction with the Library Systems Transformation programme and other strategic initiatives, the plan of work is intended to lay the foundations for the new collection metadata infrastructure necessary to realise the British Library's strategic objectives.

## The development of a consolidated Descriptive Cataloging for Rare Materials (DCRM) standard based on RDA

**Iris O'Brien, Early Printed Collections Cataloguing and Processing Manager, The British Library**

In January 2013, at the ALA Midwinter Meeting, the Rare Books and Manuscript Section Bibliographic Standards Committee (BSC) charged the DCRM(B) for RDA Revision Group to move rare materials cataloging forward in terms of harmonization with RDA. The formation of this new editorial group was recommended by the DCRM-RDA Task Force, which had submitted its final report to BSC in October 2012.

At the June 2013 ALA Annual Meeting in Chicago, the DCRM(B) for RDA Revision Group and the Bibliographic Standards Committee discussed a change of approach for developing an RDA-compatible version of DCRM. Rather than working on the DCRM(B) (Descriptive Cataloging for Rare Materials (Books)) text in isolation, the revision group proposed to work on a consolidated DCRM based on RDA, incorporating guidelines for all formats in the DCRM suite in a single text.

At, and after, the June 2014 ALA Annual Meeting in Las Vegas, the group determined that rather than a complete stand-alone version of DCRM for RDA, a task force could be formed to draft a more concise set of cataloging guidelines consisting of rules for rare materials cataloging only where deviation from RDA was necessary or where RDA required elaboration. This approach was modeled on the Library of Congress-Program for Cooperative Cataloging Policy Statements.

In 2016, the task force decided to formally name their guidelines the RBMS Policy Statements (RBMS PS), in alignment with the naming conventions of other RDA policy statements. The idea was that the RBMS PS would be an altogether new text. Although the policy statements would be informed by existing DCRM manuals, they would not be considered a revision of DCRM.

At ALA Annual 2017, the Descriptive Cataloging of Rare Materials Task Force submitted an initial draft of the RBMS Policy Statements and disbanded. The responsibility for the work of completing and publishing the RBMS PS then passed onto the RBMS Bibliographic Standards Committee. Initial community review of the draft guidelines took place in the form of online discussions of specific on the DCRM-L e-mail list from November 2017 to March 2018. However, due to the RDA Toolkit Restructure and Redesign (3R) Project taking place, which meant that all RDA Toolkit content was frozen, the policy statements couldn't be published within the RDA Toolkit as initially intended and a detailed review of the guidelines by the Rare Books Community had to be put on hold as well.

A Beta version of the new RDA Toolkit was finally released in June 2018. However, the product was far from being complete and none of the policy statements included in the full RDA Toolkit had been included in the Beta version, either. The changes were also far more substantial than first anticipated. Due to the ongoing development of the RDA Beta Toolkit, the BSC RBMS PS Sub-group decided to suspend work on the RBMS PS content until a fuller version, with a stable English text was available. However, the decision was taken that initially only policy statements for books (and possibly serials) would be released, instead of releasing RBMS PS for all formats (apart from manuscripts) covered in the original DCRM manuals as first envisaged, and to collaborate with other communities to finish the policy statements for cartographics, graphics, and music at a later stage.

Even though the latest release of the RDA Beta Toolkit (April 2019) now includes the stabilized English language version, there are still a lot of uncertainties as it is still not clear what form the policy statements in the new RDA Toolkit will take. So currently work on the RBMS PS is still suspended until the RBMS PS Sub-group has more information from ALA Publishing on this.

The RDA Steering Committee has created an RDA Policy Statement Working Group which is expected to publish some guidance on policy statements soon. At the ALA annual conference in June 2019, the future form of the RBMS Policy statements, in light of the substantial changes embodied in RDA's recent revision, was discussed at various open meetings. These discussions revealed a preference for providing a revised RDA-compatible Descriptive Cataloging of Rare Materials (DCRM) on the rbms.info website, with links to the DCRM revision from RBMS Policy Statements in the RDA Toolkit. The RBMS Bibliographic Standards Committee has therefore decided to work towards DCRM (Descriptive Cataloging of Rare Materials) as an integrated resource, with brief RBMS Policy Statements within the Toolkit that link to a more detailed, freely available document residing outside the Toolkit. However, it still remains to be seen whether this form is possible or not. In the interim, the RBMS Bibliographic Standards Committee acknowledges the need for cataloguers to continue creating records for rare materials while their institutions may have opted to switch over to cataloguing in RDA. To address that need, the committee has published a set of recommendations regarding the use of DCRM and RDA, which can be found at http://rbms.info/dcrm/rda/.

# Metadata, back to the future!

**Will Peaden, Information Resources Specialist (Metadata and Books), Aston University**

## Introduction

The future of metadata and cataloguing is a hot topic for me at present. I work alone at my University library as the sole metadata practitioner, and see hundreds or thousands of records pour into my catalogue all the time. I make global interventions to these; modifying a URL here, adding in "linkyMARC" there. I check and update authority data and keep a local authority file up-to-date. I even get to catalogue something once in a while (exciting things like a British Standard, or a Chinese language textbook!). What I don't get to do is meticulously check over the records entering our LMS, and then catalogue the items one at a time, ensuring that all the full stops are correct or that all the subject headings actually reflect the content of the work. At a certain point, given the sheer amount of supplied metadata, I have to accept what I get and make it accessible to our end users. This may mean leaving records alone and not making enhancements that I otherwise would make had I catalogued the item myself. It means trusting that the data conforms to library standards, e.g. accurate Library of Congress Subject Headings (LSCH), accurate name headings, accurate Dewey classmarks, even accurate transcription of titles (in large e-book packages, assuming the title is even supplied), and accurate ISBD.

I was therefore, excited to be invited to a Jisc "metadata day" at Senate House, in London on 22nd May. I was invited as a representative of the "Mercian metadata special interest group" part of the Mercian collaboration. The Mercian collaboration "is a professional network formed from SCONUL higher education libraries based in the East and West Midlands of the UK".[1] My task was to represent the interests of the 23 HE organisations within that collaboration. These institutions cover a wide range of HE organisations, from the Open University to very new Universities such as Bishop Grosseteste. Many of these organisations are like Aston with one metadata specialist; others have small teams (and large research collections) and some have no metadata specialists. Yet each library still faces a set of the same or similar issues, such as what titles and resources do we have access to and do we have records for these? Are there records for titles we don't have access to? Are the URLs still up-to-date, and what do we do if they need bulk changes? How do we deal with the legacy collections and metadata? And finally, most of the metadata specialists in the collaboration agree that vendor supplied meta data leaves much to be desired.

The meeting in London was conducted under Chatham House rules, so in the remaining part of this paper I will avoid relating the details of the very productive discussions held on the day. The day concluded however, with an understanding that there were two main issues around UK data supply, these being the standards for metadata quality, and licences for use and reuse of this data. The group also agreed that these issues were solvable and made a commitment to do so. What I propose to do in the rest of the paper is to reflect on the kinds of issues raised by NBK and UK data supply, as well as look at some of the standards which may emerge from these discussions.[2]

1. From the Mercian website. https://merciancollaboration.org.uk/about-mercian-collaboration accessed 3rd June 2019

2. It is worth noting that I volunteered to work on the metadata quality task and finish group.

**Metadata quality**

As already mentioned, most of the metadata at Aston is derived from 3rd parties and vendors. This data is in the form of shelf-ready MARC records for both print and electronic single purchases, OCLC WorldShare records for some electronic subscriptions, and vendor records for various other subscriptions, such as ProQuest's academic complete. Only a very few items are catalogued in house. These tend to be foreign language materials, British Standards and some audio-visual materials. The quality of the data from these sources is quite variable, from excellent to poor, with some vendors being much better than others.

**Whose standard? Which exchange format?**

Since I started in post in November 2016, Aston has technically been an RDA library. RDA is, for me, my lingua Franca for cataloguing. I was taught to catalogue in AACR2, an education I am very grateful for, but for almost my entire professional career RDA is the standard I have used and know best (notwithstanding recent changes). The promise of RDA is something I am excited by as a practitioner and librarian. The idea we could make knowledge more available and more discoverable is the reason I am a cataloguer in the first place. So, when asked by our acquisition librarian, "Are we an RDA library?" I naturally replied; "Absolutely", and therefore our suppliers need to provide an RDA record. But do 3xx fields an RDA record make? And where do FAST headings fit into this picture? Or publisher subject headings? Are the LCSH correct, accurate and reflective of the content? If RDA is more about relationships and identities, where are these in our vendor records? And then we enter the thorny issue of who owns the record rights. What changes can I make to them? Where and how can I share these records? What will the record look like in a reading list? A union catalogue? A discovery layer?

I raise these questions quite deliberately, as I do not have answers to them. I would love to know precisely what data I am getting, how I may use and reuse it and what it can do to enrich the information universe which I and our users inhabit. I would like to know that the users are able to find, identity, select, obtain and explore the resources they need or want, and that this is facilitated, made easier and possible by the metadata efforts I and my cataloguing colleagues are making each and every day. But I do not have this confidence and I do not know if the efforts I make on a day to day basis are adding the value they should.

This is not exclusively, or even primarily, a problem of derived data. Machine transcription of the primary elements of a metadata description strikes me as eminently sensible, particularly as this data already starts life digitally. Titles, authors and editors, page numbers, dates of publication and copyright, contents notes and descriptions of content are all necessary, but as a metadata specialist I'm interested in the relationships and the entities. And for those exploring the information universe this is where metadata specialists add real value. But MARC records are not really cutting it. It is true that we have relationship designators and links to authority files like NAF[3] and VIAF[4] but when we code our MARC records as RDA, what we mean is that we've added some elements and rebranded some others. As an exchange format MARC has had a long and illustrious career. However, it is no longer really required for exchange, and as an encoding schema for RDA it isn't fit for purpose. What I really want is to utilise the potential power of RDA as schema in an encoding language that understands this data. I want to be able to push the data out to systems that understand RDA, and is accessible to our users wherever they encounter it on the web.

3. The Library of Congress Name Authority File

4. OCLC's Virtual International Authority File

**The future**

At the end of the metadata day two task and finish groups were established, one to address standards and metadata quality, and another to address use and reuse of data. I have left off the discussion on reuse here but there will need to be a robust discussion around this. As we use more and more data from 3rd parties we still need to do our jobs as librarians, serving our immediate users as well as the wider community. Hopefully an open metadata licence can be created that is acceptable to both vendors and librarians. I stepped forward to help with the metadata quality issues however, as this personally interests me. What I want to try to achieve if possible is a set of standards that can be feasibly supplied by our data suppliers and which meet our needs as a library community. I do not know what this might look like at this time, but with Jisc asking the question and NBK giving some momentum to a UK standard, can we break the inertia around MARC? Can the BL, HE libraries, Jisc and Cilip come up with a UK encoding scheme for RDA that truly unlocks RDA's potential? Can we separate out recording metadata elements from the encoding and exchange format? And can we have a UK policy around the standard of data, or a UK "application profile" for RDA?

I will be working with Jisc on the metadata quality task and finish group. The success of this will depend on community engagement. What we need to know is what we want and why. To this end I really encourage anyone who thinks these issues are important to contact me or others involved in the work, with comments, suggestions, criticisms or any ideas. As far as I am concerned no idea is too bizarre or revolutionary. We are all working everyday with metadata and we are the experts. Let's try to effect the change we want to see, rather than having this change imposed on us by a world that has already moved on.

**Book Review: *Ethical Questions in Name Authority Control*, edited by Jane Sandberg. Sacramento, CA.: Library Juice Press, 2019, vii, 410 pages, ISBN 9781634000543 (paperback)**

**Reviewed by Liz Antell, Resource Assistant, Edinburgh Napier University**

*Ethical Questions in Name Authority Control* is a collection of eighteen essays which explore various ethical issues in name authority work. The volume is edited by Jane Sandberg, Electronic Resources Librarian at Linn-Benton Community College in Oregon, USA. The essays are authored by Violet B. Fox and Kelly Swickard, Michelle M. Kazmer, Jennifer M. Martin, Thomas A. Whittaker, Erin Elzi and Katherine M. Crowe, Heather K. Hughes, Kalani Adolpho, Ahava Cohen, Hale Polebaum-Freeman, Naomi Shiraishi, Travis L. Wagner, Sholeh Arastoopoor and Fatemeh Ahmadinasab, Ruth Kitchin Tillman, Anchalee Punigabutra-Roberts, Alexis A. Antracoli and Katy Rawdon, Tina Gross and Violet B. Fox, Carol Rigby and Riel Gallant, and Naomi R. Caldwell. See https://litwinbooks.com/books/ethical-questions-in-name-authority-control/ for a full table of contents listing all chapters and authors.

Discussions around ethics in cataloguing (and library and information work more broadly) are very topical at present, with a current effort underway from American and Canadian cataloguing groups along with CIG to create a set of ethical guidelines for cataloguers. There is a needed acknowledgement of the power that we wield when naming, describing, and categorising materials and, crucially for this collection, people.
The strength of this edited volume comes from its centring of diverse voices. The separate chapters provide space to explore distinct issues relating to particular groups, cultures, places, systems, and materials. At the same time, reading the chapters alongside each other reveals a number of shared concerns around language, privacy, consent, normativity, othering, and our ethical responsibilities as cataloguers, metadata workers, and information professionals of all types.

Writing about the particular areas in which they have expertise, the authors demonstrate the ways in which various standards and systems have been designed with only certain people's needs in mind and therefore fail to meet the needs of other groups. For example, many of the essays touch upon the question of the kinds of names which fit our name authority systems. Many of these systems are founded on the assumption of a name composed of two main parts, a forename and surname, both of which are usually permanent throughout a person's life. This is less likely to be the case for individuals who belong to naming traditions which have fewer or more parts to their name (which includes many non-Anglophone cultures), go by several different names, have versions of their name in different languages and scripts, or change their name over the course of their life (for example, many trans people, or women who change their surname upon marriage).

There are many implications resulting from this. Erin Elzi and Katherine M. Crowe make the point that Native American names not presented as 'Surname, Forename' will be "mangled in citations", hampering the visibility of these authors' work and so the reach and impact of their ideas. Sholeh Arastoopoor and Fatemeh Ahmadinasab in their chapter highlight that different transliterations of the same name can cause research outputs to be spread out under different headings, with negative implications for discovery and the rankings of institutions and authors affected. Alexis A. Antracoli and Katy Rawdon show that a lack of name authority work in archival descriptions impedes research access to materials relating to black history. These examples, and many others throughout the volume, illustrate the ways in which conventional naming practices in library and information work can disproportionately inconvenience, otherwise, and erase already marginalised groups.

Another key theme which arises multiple times throughout the collection is around the extent of personal data stored in a name authority record. There are multiple concerns in this area: inclusion of authors' dates of birth posing a risk to privacy, outing or misgendering transgender individuals through inclusion of gender information, othering people's identities by including personal data only when it is non-normative (e.g. recording an author's sexual orientation only when it is not straight, or an author's race only when it is not white).

Questions are raised around what sources may be deemed authoritative and accurate sources of information about a person. As Hale Polebaum-Freeman observes in their chapter, “Race and gender can’t be read off people like book dust jackets.”

Many of the essays emphasise the importance of contacting authors directly to confirm information and obtain consent to include it in name authority records. Tina Gross and Violet B. Fox argue convincingly for the important of communication with the author, not just for verification and consent but also to explain to authors the importance of name authority control work and its benefits in assisting readers to find, identify, and utilise their publications. Consent of the author is also important since there will be cases when an author does not wish to be identified either with a particular work or a particular name, and being identified could put them at risk. This is raised by Violet B. Fox and Kelly Swickard in the context of zine authors, who often include highly personal information anonymously in their works, and should not have their identities revealed without consent through cataloguers’ investigations. This echoes a call for radical empathy within cataloguing which attendees of the CIG conference 2018 will remember from Joshua Barton’s presentation, ‘Zine cataloguing, ethics & future discovery’ (published in *Catalogue & Index* Issue 193 December 2018). Other essays stress the importance of working closely with communities, not just individuals, to produce naming conventions and information retrieval systems which meet their needs, for example the multilingual and multiscript catalogue developed in Nunavut (see Carol Rigby and Riel Gallant’s chapter).

A single criticism of the book could be its structure. The chapters are grouped under five sections – Self-determination and Privacy, Impacts of Colonialism, Gender Variance and Transgender Identities, Challenges to the Digital Scholarly Record, and Emancipatory Collaborations. Given the wide-ranging content of the essays, these section designations could appear somewhat restrictive. For example, self-determination and privacy are concerns which run throughout the collection, not merely the first section. Further, a strong point made by both Kalani Adolpho and Hale Polebaum-Freeman in their chapters is the importance of attention to the cultural context of gender and the need for a deconstruction of the Western binary gender model, which has been imposed on non-Western cultures as part of colonialism. A greater level of intersectional analysis could be brought to these concerns by a framing which did not ostensibly separate gender issues from colonial issues. However, the reader is free to draw such connections between essays throughout the book, and the collocation of all these various concerns and perspectives within one volume facilitates that. As such, this is a very limited criticism.

Overall, this volume provides a thought-provoking journey through a number of important ethical issues in name authority control. Individual chapters vary in their level of technical detail but all are approachable for readers with any level of experience or interest in authority work. Each chapter stands alone and so readers are at liberty to target areas of particular interest to them, though a deeper understanding of the issues and their interrelations will be gained from an attentive reading of the full volume.

This book is available from: http://libraryjuicepress.com/

**Book Review: *Social Tagging in a Linked Data Environment,* edited by Diane Rasmussen Pennington & Louise Spiteri. London: Facet Publishing, 2019, 240 pages, ISBN 9781783303380 (paperback)**

**Reviewed by Getaneh Alemu, Cataloguing & Metadata Librarian, Solent University**

Editors of this volume Diane Rasmussen Pennington and Louise F. Spiteri introduced this book by providing a brief history of the notion of users adding descriptors, tags and reviews on the works they consult to read. This notion, as the editors explain, goes back to the 1990s when the likes of O'Connor (1996) who explored the idea of a "community memory interface" that would allow users describe the about-ness of images in a digital library. This reminds me the work of Paul Otlet who in the 1930s discussed about the importance of readers who leave trails of annotations, comments and reviews of books which he calls it "the social space of documents" (Wright, 2007). The editors as well as the contributors of this book cite the notable original works on tagging – these include Vander Wal, 2007; Smith, 2007; Mathes, 2004; Spiteri, 2006; Peters & Becker, 2009; Marinho et al., 2012; Blokdyk, 2017). They chronicle the development of the web from Web 1.0 to Web 2.0 and Web 3.0 – the Linked Data Web, also called the Semantic web.

**The case for academic research on social tagging and Linked Data**

In chapter one, the editors' indicate the years 2004 to 2007 as times when social tagging became prevalent in blogs, conferences and research articles along with its implementation by tech companies and application developers such as Twitter, YouTube,. Instagram, Google Images, Spotify and others. They also consider hashtags (#) as part of this phenomenon. The editors indicate the growing research and academic interest in social tagging and implications on traditional knowledge organisation systems and tools such as taxonomies, ontologies, information architecture, records and information management and bibliographic description. In addition, the book examines the link between social tagging and Linked Data principles – delineating a three stage development of the web, namely Web 1.0, Web 2.0 and Web 3.0. Having an extensive experience of teaching and research in the standards based library and information science field of cataloguing, classification, thesaurus construction, records management, knowledge management, information seeking behaviours, the editors are well positioned to tackle the new environments of social tagging and Linked Data environments.

**Folksonomies enriched by semantics through Linked Data**

In chapter two, Laura Cagnazzo takes a look at how the collaborative nature of web 2.0 (social tagging) yields richness and diversity of information and web 3.0 (linked data) offering the technical framework to cross-link and provide semantics. The author argues that the two approaches are complementary. Data, as Cagnazzo, argues does not only affect all of us and every aspect of our lives but we are also co-creators of data: whether that is adding/editing a wiki entry or populating a street map with geographical coordinates – such collaborative efforts gives way to a social and linked data environment. Cagnazzo reviews Tim Berners-Lee's four principles of Linked Data, that is the use of uniform resource identifiers (URIs) as names of things; use of HTTP URIs for people to look up information about these names; use of RDF and SPARQL as data formats as well as enriching the links with more links using URIs. In addition, by citing the seminal work of Berners-Lee, Hendler, & Lassila (2001) Cagnazzo outlines the underlying principles and technologies of the semantic web such as RDF, XML, OWL, SPARQL, OWL and JSON. It is also noted that Linked Data and the semantic web are extensions of the existing web. Some of the notable RDA projects and experiments such as DBpedia (apparently consisting of 3 billion RDF triples (links), OpenGraph Protocol, Google Knowledge Graph, FOAF, MusicBrainz, Zemanta, Semantically-Interlinked Online Communities Project (SIOC) – providing what is called open data graph, social or semantic graph.

Whilst Social tagging (Web 2) applications such as Delicious, Flickr and Connotea proved how the distributed and collaborative approaches could be utilised, as the author notes, they exhibit limitations of lacking semantics and context. Hence, Cagnazzo suggests the application of Linked Data. The author offers examples of applications who applied these principles such as TagOntology, Favik, TagCommons, Socal Semantic Cloud of Tags (SCOT).

Citing Choudhury, Breslin and Passant (2009), Cagnazzo explains how un-controlled and unruly metadata could be semantically and usefully linked providing a semantically-enriched folksonomy. The role of unique identification of metadata contributors as well as the metadata generated from various users is emphasised. For this to happen, the data gathered from social tagging needs to be cleaned, annotated, and disambiguated.

In Chapter three, Sue Yeon Syn describes social tagging as a user-driven collective taxonomy. The author points to the limitations of social tagging such as its inability to disambiguate homographs (words of the same spellings but may mean different things when used), collocate synonymous words, misspellings and lack of hierarchical structure that is present in taxonomies. Syn then explains how RDF and Linked Data can be used to format tags as triples and thus define semantic meaning and relationships. These requires systematic methodology and data modelling principles so as to make it easier for users to automatically create the Linked Data format of subject-predicate-object triples. The author provides examples of such applications as Revyu.com, int.ere.st, LODr, Linked Tag, SCOT (Semantic Social Cloud of Tags) and MOAT (Meaning of A Tag).

In chapter four, Ryan Deschamps talks about the impact of social tagging, in particular hashtags on Twitter affecting public policy and political decisions such as elections. In chapter five, Louise Spiteri illustrates how user-generated metadata can help enhance controlled vocabularies and enrich bibliographic records. The author argues that hashtags proved to be useful and relevant in many information discovery applications outside libraries and thus can be applied in discovery systems to allow dynamic tagging of content to highlight new or emerging topics. Once a given resource in a discovery system it tagged, this could also be linked to social applications such as Twitter, Facebook, Instagram, Pinterest and YouTube. Similarly, in Chapter Six, Laurie Bonnici and Jinxuan Ma explain how a new horizon of dynamic social interaction on social networks fostered a new avenue for social information discoverability. Users provide critical data to design interfaces and discovery systems. As Max Dobson explains, system designers could learn from online user information behaviour including their tags, information search/discovery paths, social platforms create information argues that tags are important to support findability and discoverability. Dobson acknowledges the chaotic nature of tags and the need to train users. Diane Rasmussen Pennington suggests the design and use of ontologies to organise tags into a semantic network where Linked Data principles and technologies such as RDF and URIs could be used. Sanjay Khanna also discusses the potential application of tagging in enterprise and workplace applications to build networks, identify talent and develop new knowledge. In the final chapter, Kishor John identifies six types of recommendation systems, namely collaborative filtering, content-based, graphic-based, demographic, knowledge-based and hybrid recommendation systems where each has its unique characteristics, strengths and limitations. Overall, recommender systems support users to make informed decisions when such information comes from trusted sources.

Overall, this is a well-researched, collaborative and useful book. It provides an in-depth look into the practical benefits of social tagging and recommendation systems. Resources could be described, made sense of, contextualised, tagged shared, re-used by anyone, anywhere on the web using social tagging. Social tagging is broader concept to denote the idea of users affixing content with a word or phrase of their choosing to help themselves re-find the same content or categorise it for easier navigation. On the other hand, Linked Data is the use of formal, standardised, syntactic and technical formats to structure and link data. The book offers the complementarity of both. The authors acknowledge the limitations of user tagging but offered recommendations such as the use of ontologies, Linked Data and the importance of building trust among user communities. I definitely recommend this book for people who are interested on user-generated metadata and how it can be used along with legacy and standards-based metadata approaches.

This book is available from: http://www.facetpublishing.co.uk/

Catalogue & Index is always pleased to welcome letters, emails and comments about articles in our issues, or on relevant topical themes. Please address these to the editors.

Jane Daniels' Metadata memory (Catalogue & Index 193, December 2018)) is an important contribution. It reminds us that catalogues of specific collections continue to be potential scholarly resources even when they no longer reflect the present holdings in that collection or even (especially?) when that collection no longer exists, and the holdings have been dispersed. The records are of historical and sociological value to scholars of ideas and combine with other historical and sociological data to be of great utility in numbers of areas of scholarly research. They are also of value to the much smaller (and dwindling?) group of those interested in cataloguing practice, interpretations of codes, etc., over the years. This is an initiative that all cataloguers should support.

A minor theoretical point that occurs to me lies in the fact that the difference between a catalogue and bibliography is that the former contains links to specific holdings. Absent those links, does the catalogue become a bibliography? Is a snapshot of a catalogue detailing holdings that do not exist and holdings that do a catalogue/bibliography hybrid in need of a new name? I have never cared for the term "metadata" (either in practical or linguistic terms), especially because it has come to mean an inferior form of cataloguing. Perhaps, however, this is one instance of it being useful simply because it comprehends both catalogues and bibliographies.

Michael Gorman,
January 27th, 2019

Interesting piece about ISBNs in the latest C&I (194, March 2019), but I was surprised to see the statement that the first ISBNs were generated in Britain in the 1960s. I am old enough to recall the introduction of standard book numbers, and they had just nine digits, not ten as stated, and were called just that: SBNs, not ISBNs. They became ten-digit ISBNs later, and the first digit had to be 0 (zero) so that it would make no difference to the check digit. Later still 1 came into use as well as 0, and later, of course, came 13-digit ISBNs consisting of 978 followed by the first nine digits of the ten-digit ISBN and, usually, a different check digit.

Ray Ward (Retired)
London
March 25th, 2019


**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

Philip Keates: philip.keates@kcl.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