

How does true RDA cataloguing change how we catalogue?

Jenny Wright

Bibliographic Data Services Limited, Dumfries, Scotland

E-mail address: jenny.wright@bdslive.com



Copyright © 2018 by Jenny Wright. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

New data models do influence the task of cataloguing. Differences in process and thinking arose when cataloguers moved from doing RDA cataloguing in the MARC21 format to doing true(r) RDA cataloguing in a relational database, modelled on the IFLA-LRM. The main issues, benefits, training needs and workflow solutions are described in full, and the effect on the task of cataloguing of using a nonMARC21 data entry interface in a relational database is compared with creating bibliographic records in the MARC21 format.

Keywords: cataloguing IFLA-LRM RDA metadata bibliographic

Introduction

With the aim of having a new database which would support better working practices and be in alignment with current concepts in cataloguing, BDS has created a new inhouse data entry system, called Aura, which is not based on the MARC21 bibliographic format. There were three principal drivers for change: the previous data entry system used outmoded technology; achieving better data management was not attainable without change; and new metadata standards were available.

The IFLA-LRM¹ conceptual model was used as the basis of the new database design. The use of entities and their relationships to organise bibliographic data allowed the reuse of cataloguing effort, and enabled flexibility in terms of the bibliographic carrier used.

As a result, cataloguers have changed the way they think about their tasks. BDS implemented the Resource Description and Access (RDA) cataloguing standard² in January 2013 and so in theory have been cataloguing using RDA for five years already. In practice, staff found that the MARC21 based data entry format kept many of the RDA concepts implicit, and therefore hidden and forgettable.

Thinking in terms of the RDA entities and their relationships has now become key and this change has required an intellectual engagement from staff, with challenges for managing change and meeting training needs.

Business context

BDS is a cataloguing agency based in Scotland, established in 1995, supplying bibliographic records to library suppliers, libraries in various sectors, and publishers. Our business is primarily in the UK, but we also have international customers seeking metadata for English language publications. BDS is the contracted provider of cataloguing-in-publication (CIP) data to the British Library. We produce in the region of 150,000 CIP and about 60,000 confirmed MARC21 records per annum.

We receive ONIX³ data from publishers, as well as some print catalogues, and we use that information to create CIP bibliographic records. We also receive copies of published resources via our customers, for which we provide full, confirmed records. Many CIP records never get upgraded to confirmed records because our customers don't buy every resource that gets published.

The basis for the original data entry interface dated from the company's inception in 1995, and whilst it still worked well it was getting harder to maintain the technology, and there were ideas for development which it could not support.

Given the proliferation of formats that many titles have, we knew that the confirmed data for one manifestation could inform another manifestation's bibliographic record, but we had no effective way to accomplish that with our MARC-based, flatfile dataset. The basis of RDA being a conceptual model of entities and relationships, i.e. IFLA-LRM, gave us an intellectual basis to improve the quality of metadata by using established data about some entities to benefit related entities.

Aura, the new inhouse cataloguing system, contains work, expression and manifestation records in a relational database, and uses the WEM data to produce MARC21 records for every manifestation. By holding data in identified, specific fields, we will be able to export bibliographic data in whichever format our customers deem useful in the future.

The process of creating Aura was interesting, but the focus of this paper is the experience of cataloguers using the new interface.

Change

Using Aura was initially bound up with the inherent nature of experiencing change, with both positive and negative elements. Staff were prompted to question many practices, and contribute to policy decisions, which promoted engagement, but it was challenging to accommodate the differences of a new methodology and interface.

Having to enter data in MARC21 requires substantial training in MARC21 coding, and ISBD punctuation, as well as the training in the descriptive cataloguing and subject access provision. So whilst it was a positive for BDS that this would no longer be necessary for trainees, it was a significant negative for existing staff that their expertise in the MARC21 coding, and ISBD punctuation, previously required to do their work was no longer required

within the data entry system. Knowledge and understanding of the MARC21 standard is still useful to the business, since that's how our data is packaged for the vast majority of our customers, and so that expertise is still useful – it's just not critical for cataloguers to be able to contribute to the database.

Expertise in MARC21 does not translate into expertise in entering data in the new, unfamiliar interface, and there was concern that some data elements would be accidentally forgotten. Reassurance is provided in the form of a “MARC21 view” option for each manifestation, which shows how the data appears in the MARC21 bibliographic record being sent to customers. This was a useful training tool, illustrating to staff how the data in Aura relates to their knowledge base.

Positive factors included no longer having to make corrections because of typographical errors or omissions in any aspect of the MARC21 coding, or the ISBD punctuation. Also, the new user interface brings together similar data elements that are often separated in the MARC21 format, for example subject classification and subject headings, which cataloguers found very helpful and promoted good practice in workflow.

However, the change that is more interesting is how the task of cataloguing is altered by using a relational database structure, focusing on the WEM entities and their relationships.

WEM records

Instead of creating a MARC21 record for each manifestation, Aura requires a work record (*Work*), expression record (*Expression*), and manifestation record (*Manifestation*) for each resource. Some resources will be the single manifestation of the single expression of the work, and others will be one manifestation among many. For example, the BDS database has over 100 manifestations for the individual work created by Jane Austen, known as “Pride and prejudice”.

New manifestation, new expression, or new work?

Copy cataloguing is a term used when one can take an existing MARC21 record and use it as the basis of a new MARC21 record. When we're using *Works*, *Expressions* and *Manifestations* the task becomes one of finding the existing *Work*, establishing whether it's a match, proceeding through the WEM entities, and creating new entities where needed. The tasks are broadly similar –starting by searching for important work information such as the name of the creator, and the title by which it is known - but our new methodology is more explicit and therefore requires greater confidence in one's ability to clearly identify and distinguish works, expressions and manifestations.

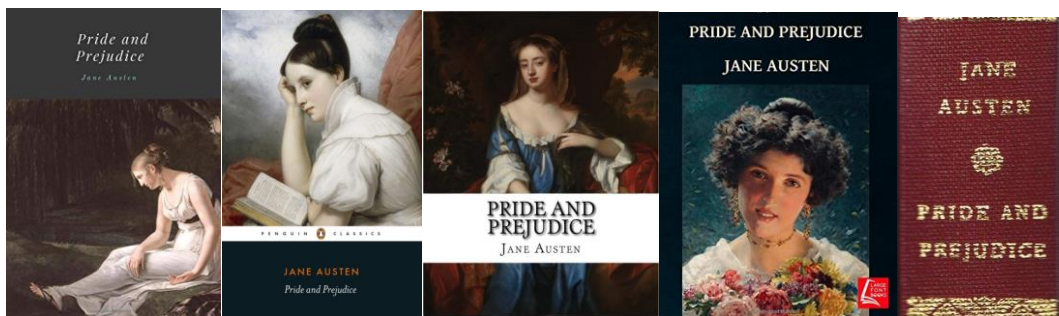
We found staff were confident in identifying new manifestations, as that's largely ISBN based and presented no great change from previous practice. New works were also generally straightforward, with “distinct creations” being understood. However, deciding if a resource represented a new expression was harder for several reasons.

Often, publishers don't provide enough information at the CIP stage for staff to have confidence in which expression, or version, of a work they are cataloguing. However, even with the resource in-hand, cataloguers were still hesitant. If the content type was different,

then that was straightforward and clear, but confidence in distinguishing between different realisations of a work required training.

Bringing a phrase from the scope notes in the IFLA-LRM to the foreground proved most helpful for staff: “*An expression comes into existence simultaneously with the creation of its first manifestation*”.

With a focus on the publication history, which was a familiar concept, cataloguers were able to rationalise whether or not they considered the resource to be a different realisation of the work. For example, these five publications of *Pride and prejudice* are all English text, with no immediately apparent distinction:



However, a look at the publication history on the title page verso provides useful information for three of them:

First published 1813
Published in Penguin Classics 1996
This edition reissued with new Chronology, updated Further Reading
and 1972 Penguin Classics Introduction by Tony Tanner 2003
10

Introduction and Notes copyright © Vivien Jones, 1996, 2003
Textual Adviser's Note and Chronology copyright © Claire Lamont, 1995, 2003
Appendix: Original Penguin Classics Introduction copyright © Tony Tanner, 1972
All rights reserved

First published, in 1993, by Wordsworth Editions Limited
8B East Street, Ware, Hertfordshire SG12 9ET
New introduction and notes added in 1999
Illustrations added in 2007

This edition is selected from the definitive text of Professor R. W. Chapman, by kind permission of the Oxford University Press. The only alterations are of a minor typographical nature, in particular the expansion of contractions, the use of single quotation marks, the reduction of excess capitalization and some clarification of the original erratic punctuation.

First Folio Society edition 1962: reset 1975
Fifteenth printing 1996

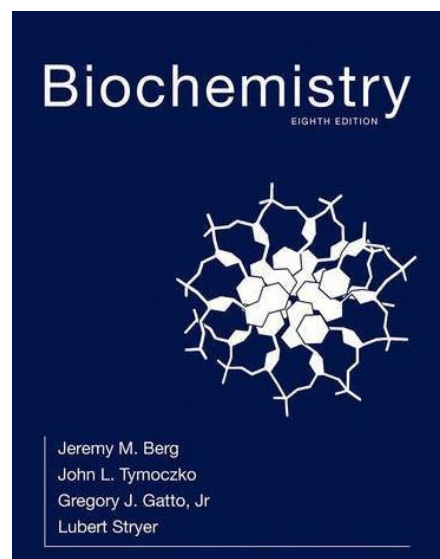
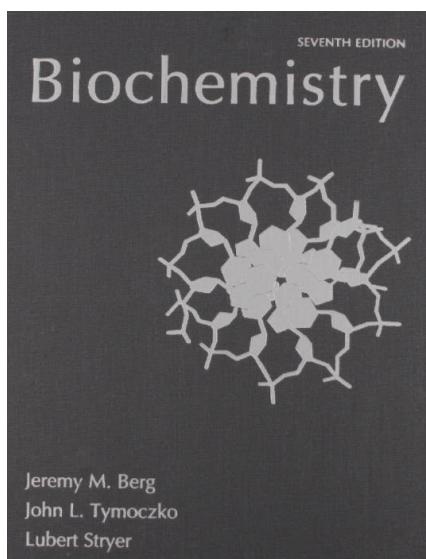
Set in 'Monotype' Fournier
and printed by Butler & Tanner Ltd, Frome
on Ibis Wove paper.
Bound by Butler & Tanner Ltd
in WW Light buckram and printed paper sides
designed by the artist

As well as the differing dates of publication providing information as to the dates of expression, we can also recognise that there are contributors to the expressions. Differing contributors indicate the need for a new *Expression*, though when the contributors are the same we cannot assume that it is the same expression.

There are some resources, usually college textbooks, which change their attributed authors over time. This presents another challenge, as in many regards it seems to the cataloguer to be the same work: the intent of the work has not changed, and the MARC21 record would certainly have been used in the copy cataloguing task. However again the IFLA-LRM scope notes provide the necessary guidance:

“For example, two textbooks both presenting an introduction to calculus, or two oil paintings of the same view (even if painted by the same artist), would be considered distinct works if independent intellectual or artistic effort was involved in their creation.”

For example:



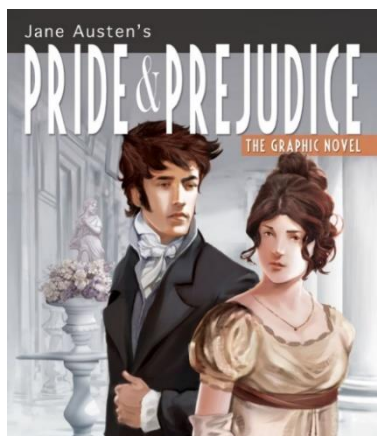
The author team has changed between the seventh and eighth editions of Biochemistry. Dr Gatto's presence is at first glance innocuous, but in fact signifies the requirement to create a new **Work**. This did not sit well with cataloguers until they realised that one cannot just add a new author to the existing **Work** without stating that they're responsible for all expressions of that work, and neither can Dr Gatto be added as a contributor to a new **Expression** of the existing work, as it is clear that he is part of the author team.

Training therefore focussed on pragmatic aspects of information sources and relationships between entities, to decide whether matching **Works**, and creating new **Expressions**, or creating new **Works** was most appropriate.

Related works

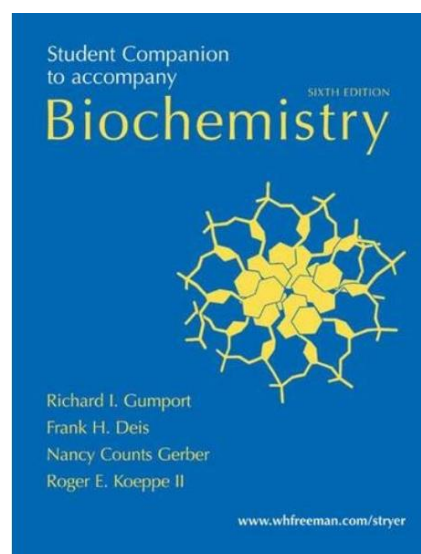
In the MARC-based system, staff knew to add an access point when our resource had some relationship to another work, or expression. However, the decision making for the access point was MARC21 based, i.e. whether it was a name/title access point (700\$t), or a title access point (730\$a), and whether the second indicator should be blank or 2. It was very new to ask them to specify the relationship. Although this is technically possible in MARC21, using the \$i subfield, it wasn't something we'd had any demand for.

Cataloguers now have to consider what the relationship is, and have a range of possibilities, including "is abridgment of (work)", "is graphic novelization of (work)", and "is parody of (work)" to choose from. We found that whilst a measure of training was required, the bigger issue was a lack of information regarding the resource, especially at the CIP stage, so using the broad relationship "has related work (work)" was often necessary. It is also worth noting that when choosing from a controlled vocabulary, Aura has the facility to show only those relationships identified as most likely, or can switch to seeing the entire list. This makes it easier and quicker for staff to choose a relationship, as they can see all of the most likely relationships without having to scroll through the entire list.



As well as having to specify the relationship, there is the additional issue of knowing whether the relationship is to another work, or to another expression. The latter is less common, but there are study guides to specific editions of texts, for example:

In training, we emphasised that when it mattered what the language, contributor, or exact content of the related entity was, then the relationship was to the expression, not just the work, and provided guidance examples to support cataloguers.



Aggregates

In a similar vein, being more stringent about the WEM entities involved in a resource means that omnibuses often require us to create **Works**, in order to be able to specify the relationship between the entities.

ANTON CHEKHOV

Plays

Ivanov
The Seagull
Uncle Vanya
Three Sisters
The Cherry Orchard

Translated with Notes by PETER CARSON
With an Introduction by RICHARD GILMAN

This is the title page for a selection of plays by Chekhov, and we can immediately see that there will be five related works, and their expressions. In the past, a contents note and the necessary name/title access points in the MARC21 record would have sufficed, but in our relational database we are required to create basic **Works** for each contained work, if they are not already in the database. We can't state a relationship to another entity, unless that entity is also in the database.

It can therefore be fairly time consuming, particularly for translated resources like this one, which need research to establish the work titles. However, those entities are then available in the database for future use. In order to keep the process as streamlined as possible, when creating the basic **Works** the only fields required are title, first creator (if applicable), audience and form.

Form of work

The RDA element Form of work is “*a core element when needed to differentiate a work from another work with the same title*”, and has no recommended controlled vocabulary. It became useful to our database to record it for every **Work** because it helps determine some coding in the MARC21 record for any manifestation, and because it helps set appropriate defaults in other parts of the **Work**, which reduce errors and increase speed of data entry. For instance, if we identify the form as Autobiography, then the system can both set the appropriate 008/34 code and check that the work has the same name heading as both creator and subject. If there's an anomaly, the user will get either a query, or a refusal, depending on the nature of the problem. We are using an inhouse list, based on the Library of Congress Genre/Form Terms⁴.

Self-referencing works

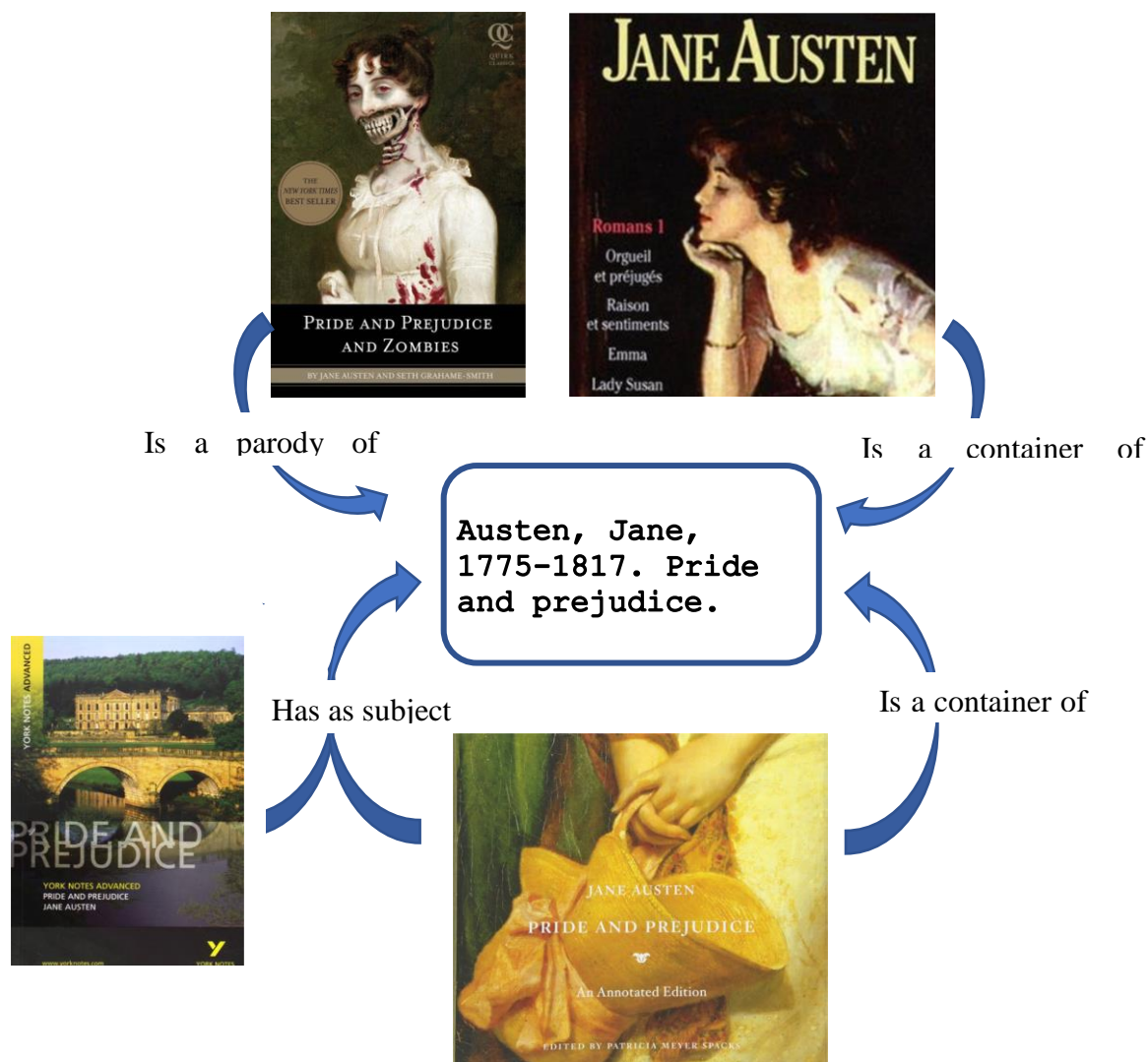
The examination of related works, aggregates, and the issues involved with establishing the correct entities and relationships, led to the realisation that we needed to educate users in the phenomenon of self-referencing entities. A familiar concept for programmers, the self-referencing entity was a new notion for our cataloguers. Linguistically, one can realise that self-referential means “refers to itself”, but in database terms we had to make sure users understood the paradoxes that would arise if a Work had itself as a subject or as a related Work. For example, if one takes the resource

Austen, Jane, 1775-1817. *Pride and prejudice*.

Annotated edition / edited by Patricia Meyer Spacks.

Text. English. 2010.

If this resource was catalogued as being the work with creator Jane Austen and title *Pride and prejudice*, and was also provided with the relationship “Has as subject” to that work, i.e. was self-referencing, then we’d end up in the situation that all expressions of that work would also have that relationship – and we know that there are many plaintext expressions of the work which do not have commentary or other critical content. This paradox is prevented by avoiding self-reference and creating a new *Work*, which both contains Jane Austen’s *Pride and prejudice* and has it as a subject.



Cataloguers are therefore confronted with the nature of aggregate works far more explicitly than they were in the MARC21 record. Even if the part work has no separate bibliographic identity, as in our example of the commentary provided by Patricia Meyer Spacks, it is intellectual content which is not part of the work known as Jane Austen’s *Pride and prejudice*.

Authority

The use of the Name Authority Records (NARs) created by the Name Authority Cooperative Program (NACO) is established practice. However, it was less obvious to cataloguers that use of preferred titles meant they were also doing authority tasks with respect to works. Having a separate *Work* and *Manifestation* in Aura made the task explicit. The MARC21 practice of having a 245 field for the title, and only a 130/240 if the work title is different from the manifestation title, made the work title implicit. When we require titles for the *Work* and for the *Manifestation*, which may or may not be the same, there is a different emphasis and Aura's separation of the entities therefore resulted in cataloguers having a far clearer understanding that they were identifying, distinguishing, and establishing *Works* as authority records.

Workflow

As indicated, the changed way of approaching the cataloguing of resources that Aura provided, has also had an impact on workflow. In the future, our database will have many established *Works*, *Expressions* and *Manifestations*. Initially, there are few and so cataloguers are importing MARC21 records as required, case by case, to populate the database with *Works* and *Expressions* as needed for new publications.

| MARC21 Database | | "Aura" WEM Database | |
|--|---|--|--|
| Receive details of new publication | | | |
| Check for possibility of previous publication of same work, for copy cataloguing | | | |
| Create new MARC21 record | Clone the previous MARC21 record, changing relevant details | Import data from publisher ONIX and edit to create new W, E, M | Import previous publication record from MARC21 database, to create W and E for reuse |
| | | | Import data from publisher ONIX and match to existing W, check whether match or need new E, edit new M |
| Export CIP MARC21 record for Manifestation | | | |
| Receive Published resource | | | |
| Update transcribed fields | | Confirm the Work | ?W may be confirmed already, depending on status of record it was created from |
| Check the transcription, cataloguing decisions and subject analysis | | Confirm the Expression | ?E may be confirmed already, depending on status of record it was created from |
| | | Update and check the Manifestation | |
| Export confirmed MARC21 record for Manifestation | | | |

There are a couple of extra steps involved compared to copy cataloguing in the MARC21 database, but they bring the reassurance of better quality metadata. Cataloguers know that in copy cataloguing it is imperative to check every detail; we also know that unfortunately data which does not apply to the new publication does sometimes slip through, or an old authority heading may not get updated as it should.

The separation of the different entities in Aura means that we can have verified, confirmed status *Works* and *Expressions*, even when the *Manifestation* is based only on publisher's advance data. It also means, and this is a significant benefit to our metadata, that if we only

see one manifestation of a work then the confirmed *Work* and *Expression* can benefit the bibliographic records for all of the other *Manifestations* we may have, though we may never see those other publications.

There is therefore less rechecking involved, due to the better provenance we can give to the separate entities. It is also the case that, in a relational database, when an authority heading is updated then any related entity is also updated. For example: the author and scientist Stephen Hawking died recently, and his authorised name heading in NACO⁵ now includes that information, having been updated:

Hawking, Stephen, 1942-2018

In a relational database, any entity with a relationship (has author, has subject) to this name heading immediately has the new information included.

We did have to take care with certain aspects of the data, so that we didn't overwrite transcribed fields for confirmed records. For instance if the publisher changes their address, we want to update the database but not overwrite those confirmed records for which the old address was correctly transcribed. The publisher table is used for a controlled entry at the CIP stage, and additional, uncontrolled fields are used for the transcribed publisher statement at the confirmed stage.

Training

As well as supporting staff by providing clear definitions, examples and policy guidance for the situations already described, a significant training issue for this new system is getting users to realise the effect that changing information in just one field can have. In the past, it was possible for a cataloguer to change the 245 field in a MARC21 record and send out one updated MARC21 record. In Aura, if someone changes the *Work* title then it will result in as many updated MARC21 records as there are *Manifestations* of all the *Expressions* of that *Work*.

Of course, some changes need to be made and to avoid paralysis there are different levels of user permissions so that an inexperienced cataloguer cannot perform unwanted changes.

Conclusion

Whilst the intellectual task of cataloguing is not intentionally changed by entering the data in a different interface, it is clear that Aura has had an impact on the way in which the descriptive and access related tasks are performed by cataloguers.

The more rigorous approach to the reuse of previous cataloguing effort which is achieved by using the WEM entities and their relationships, based on the IFLA-LRM conceptual model, and the improved understanding of the WEM entities that results from treating them as separate, albeit related, entities, has resulted in the task of cataloguing being a more intellectually satisfying process. When we have sufficient records in Aura, in time, I hope to be able to demonstrate quantifiable improvements to the quality of the metadata.

Acknowledgments

Thanks to BDS for their support, and to those publishers I've used images from.

References

1. Pat Riva, Patrick Le Bœuf, and Maja Žumer, 2017. IFLA Library Reference Model A Conceptual Model for Bibliographic Information
https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf
2. RDA Toolkit <https://access.rdatoolkit.org/>
3. ONIX standard <http://www.editeur.org/8/ONIX/>
4. Library of Congress Genre/Form terms
<https://www.loc.gov/catdir/cpsd/genreformgeneral.html>
5. Stephen Hawking's NACO heading <https://lcn.loc.gov/n81020731>