Welcome to C&I 177, which features papers from this year’s conference held at University of Kent, Canterbury, on 8-10 September. Speaking as a delegate, I found the conference to be both enjoyable and highly relevant and I wish to thank the speakers who have contributed papers for this issue. The conference was organised into four themes: the impact of metadata standards, the impact on the organisation, the impact of metadata on users and the impact of metadata professionals, and we have papers representing each theme. It should be interesting to both those who attended the conference and those unable to make it.

The next issue of C&I, number 178, will be an "open mike" issue, so there has never been a better time to write for C&I! Just contact the editors with your suggestion.

With this issue we say goodbye to co-editor Heather Jardine. Heather has done a fine job of co-editing C&I since June 2011 (issue 163). On behalf of CIG, I want to thank Heather for all her hard work and expertise in putting the journal together in this time and for helping me to get started as co-editor. We are pleased to announce that Karen Pierce, Cardiff University, will be taking over the role of co-editor in 2015.

Helen Garner (h.j.garner@shu.ac.uk)

Karen Pierce (PierceKF@Cardiff.ac.uk)
Introduction

Following the replacement of AACR2 by RDA, BIBFRAME is widely viewed as the replacement for MARC. Much like MARC, it was initiated by the Library of Congress. BIBFRAME is an abbreviation - not an acronym despite the capitalisation - for the BIBliographic FRAMEwork Initiative. It is also frequently written as “Bibframe” which this article will use from now on. Bibframe’s initial remit was wider than establishing a technical standard, although that is the current major focus of its work and of this article. This article will try to assess what Bibframe’s impact is likely to be, especially on cataloguing.

The Bibframe initiative was initially undertaken with the consultants Zepheira - whose president Eric Miller was involved with the RDF specification itself - in partial response to RDA testing in 2011 which determined that MARC21 wasn’t up to handling RDA properly:

“Most felt any benefits of RDA would be largely unrealized in a MARC environment. MARC may hinder the separation of elements and ability to use URIs in a linked data environment.”

For RDA to be adopted, the Committee suggested that the national libraries

"Demonstrate credible progress towards a replacement for MARC."

The Library of Congress decided in 2011 to use linked data as the basis of a replacement:

“The new bibliographic framework project will be focused on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model.”

This was a highly significant decision, both for determining the shape of Bibframe and the context in which it is to be placed.

Linked Data

Linked data is not a tightly defined technical standard but an approach and set of technologies that aim to bring the benefits of the web to data, not just to documents. This is how Tim Berners-Lee described it in 2006:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs so that they can discover more things.5

URIs are basically URLs used as identifiers. In cataloguing we commonly use authorised strings as identifiers. These change when life events or rules change, but URIs do not: they are more like system numbers but are not confined to particular databases and are universal. HTTP URIs are simply URIs on the web. When a human looks up a URI they should get a textual document in HTML to read; when a computer looks up the same URI, it will prefer instead some data in RDF, which provides a way of representing data on the web. What made the web good was linking to other things, and following your nose to find out more. The same is also the case for linked data. In cataloguing terms, you don't have to provide all the information about an author in the record for a book if you can follow a link to an authority record or to a page about the author.

In short, linked data gives us a web of data rather than a web of documents, and it is RDF that gives linked data its basic shape.

RDF

RDF is a data model based (as is FRBR) on the entity-relationship model. All facts or assertions are expressed as triples. An assertion such as:

**Brideshead revisited** was written by **Evelyn Waugh**

can be broken down into two entities ("**Brideshead revisited**" and "**Evelyn Waugh**") and a relationship ("**was written by**" or "**creator**"):

Rather than using text to identify the book in this case, we can use a URI to unambiguously identify which book we are talking about. The Library of Congress has published a URI for Brideshead Revisited:

http://id.loc.gov/authorities/names/no97080492

creator

Evelyn Waugh

Similarly, LC have a URI for Evelyn Waugh, and Dublin Core have a URI for the creator relationship itself:

http://purl.org/dc/terms/creator

http://id.loc.gov/authorities/names/n79049248
This is a triple expressed graphically. We could also express this as text, with some abbreviations to make it readable:

@prefix lcn: <http://id.loc.gov/authorities/names/>.
@prefix dcterms: <http://purl.org/dc/terms/>.

lcn:no97080492  dcterms:creator  lcn:n79049248.

This way of writing RDF is called turtle. There are several such ways to write RDF, the most common one being RDF/XML, although this is harder for people to read. The BNB (British National Bibliography) has released a large amount of linked data. Here is a real example based on the way the BNB would say the same thing, first as a graphic...

... and as turtle:

@prefix bnbr: <http://bnb.data.bl.uk/doc/resource/>.
@prefix bnbp: <http://bnb.data.bl.uk/id/person/>.
@prefix dcterms: <http://purl.org/dc/terms/>.


The Impact of Using Linked Data

Using linked data sends a number of powerful messages. First, Tim Berners-Lee thinks it is a good idea! Linked data builds upon the incredible success of the web and extends it from documents to data.

Second, it is not a cataloguing standard, nor even a library standard, but a web standard. RDF and HTTP for instance are W3C-maintained standards. This gives libraries and catalogues the opportunity to escape from many of the silos we have complained about being trapped in. By contrast, MARC is firmly a cataloguing standard and arguably not even a library-wide standard: libraries are full of repositories, archives, and databases using anything but MARC. MARC also requires highly specialised software to use it and makes it difficult to share our data or include others’ data.

Third, linked data doesn’t require but often assumes a notion of openness wholly missing from MARC. Most MARC records are essentially unlicensed where the assumption must be that they are copyrighted. By releasing data openly under a specific licence, people can re-use or re-purpose our data. It fits in really well with the open access, open software, and open data movements.

A number of libraries have already published linked bibliographic data. The following is merely a selection⁷:

- 2008  Swedish National Library⁸
- 2011  BNB⁹
  - Cambridge University Library¹⁰
  - Europeana¹¹
  - French National Library¹²

---

7. The list doesn’t include a number of authority files such as VIAF (OCLC’s Virtual Authority File), the Library of Congress (including its Name Authorities and LCSH), and the German National Library.


2012  OCLC Worldcat (using schema.org)\textsuperscript{13}
Spanish National Library\textsuperscript{14}
2014  RLUK (as part of the European Library)\textsuperscript{15}

These libraries have not only already established models but have actually published data. It is also worth noting that these models all differ, using different properties from different vocabularies. This makes sense if the needs and priorities of these libraries are not homogenous.

Bibframe as RDF

So what does Bibframe actually look like? Above we looked at an example from the BNB. Below we can see more statements made about the author by the BNB:

Note that three vocabularies have been used by the BNB for various properties here: Dublin Core Terms\textsuperscript{16} (dct), RDF Schema\textsuperscript{17} (rdfs), and OWL, the Web Ontology Language\textsuperscript{18} (owl).

\begin{itemize}
\item 15. RLUK. RLUK Hack Day #RLUKhack. 2014. http://www.rluk.ac.uk/events/rluk-hack-day/
\item 17. OWL. RDF Schema 1.1. 2014. http://www.w3.org/TR/rdf-schema/
\end{itemize}
Here is the Bibframe equivalent of the triple, asserting Waugh to be the author of Brideshead Revisited. It uses an imaginary implementation at UCL:

![Bibframe diagram](image)

There is a Bibframe-specific version of the creator property with a locally authorised author. Here is some more detail about that author:

![Bibframe diagram](image)

The thing to note here is that both additional properties - bf:authorizedAccessPoint and bf:hasAuthority - are also Bibframe-specific. This is significant. To get a better appreciation of this, let’s look at the description for a whole book, the equivalent of a MARC record:
This is data for a book on the BNB. It is unreadable squashed up like this but you can see the variety of vocabularies used as they are in different colours. There are ten and it looks fairly varied. The lines in red are the only ones that use terms made up by the British Library as they explicitly attempted to re-use existing vocabularies, as did the Bibliotheque nationale de France for their data:

“We preferred to reuse existing vocabularies in order to foster interoperability.”

Below is data for the same book taken from The European Library (TEL):
The red in this case represents RDA Element vocabularies. There is a solid block of it as it is useful for describing books, but there is also a fair sprinkling of other vocabularies. The Europeana Data Model, the nearest thing to a bespoke vocabulary is the first white one at the top mentioned in the prefixes, although it’s not actually used in this data at all!

Below is the data for the same book using the Bibframe utility that converts data from LC catalogue records:

Note how monochrome it is by comparison with the two above! The red is the Bibframe vocabulary, orange is rdfs, blue is madsrdf (used for authorities). This has been a strong early criticism of Bibframe: that linked data practice generally welcomed re-using vocabularies but Bibframe has attempted not to. Bibframe however, has brought forward several reasons for its approach, which can be summarised as authority and stability. A 2012

---


report from the Library of Congress said that:

“While the recommendation of a singular namespace is counter to several current Linked Data bibliographic efforts, it is crucial to clarify responsibility and authority behind the schematic framework of BIBFRAME in order to minimize confusion and reduce the complexity of the resulting data formats.”\(^{23}\)

Although it is hard to see that maintaining a single data format for all bibliographic purposes will necessarily reduce complexity!

Relying on others is also potentially risky for stability. If libraries start using Dublin Core and it disappears, or the defined meanings change, then what? Arguably though, URIs still retain their meaning by consent even if there is no content on the web to back them up. The Bibframe approach also has the opposite flaw of putting all its eggs in one basket: if Bibframe fails then more would be lost than with a single smaller vocabulary. It is perhaps instructive to consider that many libraries used - and still use - UKMARC despite its official demise some years ago. The Bibframe plan is instead to stabilize the vocabulary then make mapping links as appropriate to other vocabularies.

Below is the common image showing the Bibframe model which demonstrates one of its most unsettling aspects for cataloguers: that it doesn’t follow FRBR!

---

The Work looks like a FRBR Work; the Instance looks very like a FRBR Manifestation; the Authority looks like any of the group 2 or 3 entities: authors, subjects, etc.; but where is the Expression? In reality, a Bibframe work can be both a FRBR work and a FRBR Expression, depending on how it describes itself:

![Diagram showing the relationships between Work, Expression, Manifestation, Instance, Item, Annotation]

This though does make the point that Bibframe is designed to handle RDA but also lots of things that are not RDA. Most AACR2 records don’t have Expression records either and there are other FRBR-like models (e.g. CIDOC CRM\(^{24}\)) to take account of if Bibframe is to move beyond purely accommodating library catalogue data.

### Maintenance of Bibframe

Bibframe is an initiative of the Library of Congress. There is as yet no governing committee for Bibframe although I understand\(^{25}\) that this will be widened out to a similar governance system to that of MARC.\(^{26}\) This is arguably a good thing as it has at least got off the ground in the absence of another clear mechanism; it means Bibframe is not in commercial hands; and it should have a strong central direction. On the other hand, a single large national library, albeit in a pivotal role, is not necessarily representative of the libraries that will want to use it. Some of the partners and early experimenters involved in Bibframe- e.g. George Washington University, the German National Library, and OCLC\(^{27}\)– do have a wider field of interest, but do not have

---


ownership as such. Consultation via the mailing list\textsuperscript{28} has been increasingly vibrant, especially in response to some specific criticisms, but decision making is still in the hands of the Library of Congress. Not involving commercial entities from the beginning is also potentially risky as their input will be vital to any scheme’s adoption in practice, especially with the centralised standards environment we currently work in.

The Purpose of Bibframe

The library linked data efforts mentioned above - e.g. those of the BNB or Swedish National Library - were instigated by those libraries for their own purposes. They drew on others’ work, but did them for their own business reasons, even if those reasons were experimental; Bibframe by contrast is interesting in being undertaken by one national library for the benefit of the community for the broadest of purposes:

“Our BIBFRAME provides a foundation for the future of bibliographic description, both on the web, and in the broader networked world.”\textsuperscript{29}

MARC, Bibframe’s predecessor, was a very similar creation by the Library of Congress but with a much narrower scope, which was testing “the feasibility of distributing Library of Congress cataloguing in machine-readable form to a variety of users” although it was recognised that “devising a method of recording bibliographical information in machine-readable form was basic to the solution of other problems.” \textsuperscript{30}

Given the low degree of automation at the time, MARC’s success was remarkable; given the massively entrenched infrastructure of thousands of libraries, differing LMS’s, vendors, and cooperative sharing schemes in existence now, the task facing Bibframe is daunting from the opposite point of view: the Library of Congress, basically single-handedly, is seeking to change an entire ecosystem and market on behalf of that ecosystem and market!

Bibframe in a Wider Environment

Before RDA, the typical stack of cataloguing standards looked like this:

\begin{itemize}
  \item Archives of BIBFRAME@LISTSERVE.LOC.GOV. http://listserv.loc.gov/listarch/bibframe.html
\end{itemize}
ISBD informs the AACR2 rules which are encoded in MARC. With the introduction of RDA and the replacement of MARC by Bibframe, it is often assumed that the new stack will look like this:

```
<table>
<thead>
<tr>
<th>ISBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AACR2</td>
</tr>
<tr>
<td>MARC</td>
</tr>
<tr>
<td>FRBR</td>
</tr>
<tr>
<td>RDA</td>
</tr>
<tr>
<td>BIBFRAME</td>
</tr>
</tbody>
</table>
```

There are however some problems with this approach. First, linked data is greater than Bibframe.

**Linked data > BIBFRAME**

As discussed above, it’s already in use in an increasing number of libraries, in some cases for current workflow, e.g. Oslo Public Library or the Swedish National Library. There are already other ways of cataloguing in linked data. As well as the vocabularies referred to above and the models used by the BNB, TEL, BnF, and others, even the whole RDA element set is available as RDF. Furthermore, it is being used: the TEL uses it for core bibliographic data.

A number of non-library organisations are adopting linked data, such as the BBC\(^\text{31}\), DBpedia (a linked data version of Wikipedia)\(^\text{32}\), the UK Government\(^\text{33}\), and Ordnance Survey\(^\text{34}\), Schema.org\(^\text{35}\), used by OCLC on its

---


32. DBpedia. http://dbpedia.org/About


Worldcat service\textsuperscript{36}, is an initiative started by search engines to improve results and embed metadata within web pages. Libraries can benefit by creating and exploiting links with these efforts and enriching our own data, allowing services to, for example, crosswalk from Wikipedia to catalogue searches via OCLC’s VIAF service\textsuperscript{37}. Both libraries and the wider world are to some extent already occupying the same space and performing the tasks that Bibframe seeks to do. Even if successful, how long will Bibframe take to reach maturity? It has gone from a recommendation in 2011 to a more stable but still very experimental vocabulary in 2014 but it is still some way off from being settled. The practical effects of this are clearly laid out by the Oslo Public Library:

“So why don’t we wait for the library standards for linked data cataloguing that are bound to come sooner or later? Well, first of all there is the suspicion that «sooner» might be slightly less likely than «later». The new Oslo Public Library is opening in only four years, and we simply don’t have the time to wait and see what happens in the meantime.”\textsuperscript{38}

Bibframe’s success depends heavily on the software it is used on, and the software culture it will exist within. It requires complete re-writing of library software if we continue to use the classic LMS. There is little in the way of clear commitment by LMS vendors to this linked data generally, and Bibframe in particular. Indeed, one concern would be an LMS that handles Bibframe but not linked data generally, whereas linked data does give us the chance to break free of library-specific software for our metadata.

**Metadata > Cataloguing**

Archival and repository metadata, along with museum and commercial metadata all use various standards - by no means all of them will ever use RDA - but they will all benefit from being linked to each other and to a wider web of data. Whether Bibframe suits the needs of all traditional catalogues is one thing; its adoption for all bibliographic purposes is quite another. Linked data at least offers a way of bridging between bibliographic (and non-bibliographic) efforts across the web. Ultimately, a future involving library linked cataloguing data might look more like this:

\textsuperscript{36} OCLC. OCLC Adds Linked Data to WorldCat.org. 20 June 2012. https://oclc.org/news/releases/2012/201238.en.html


There is a variety not only of vocabularies and models, but of rules and initiatives feeding the data. Whatever works, whatever fits libraries’ various business cases, and fulfils the varied needs of our varied users will win out. Is this likely to be one scheme under the ultimate direction of one library? Bibframe, like RDA, will certainly be very influential because of its pedigree, who is doing it, and the strong need it is trying to fulfil, but those three things could also hold it back if it is not agile enough, does not join in properly the with the rest of the web of data, and does not look to satisfying newer and broader aims beyond the traditional catalogue.
OU Digital Archive & Metadata

The OU Digital Archive (OUDA) has been developed by the OU’s Library Services to preserve, manage and provide access to digital and digitized heritage content from the University Archive. OUDA includes study materials, videos, audio recordings, images and other digital materials selected from the University’s 45 year history.

OUDA use at least 11 schemas & 35 different vocabularies. These standards are used in 12 profiles: AV Descriptive, AV Digital, AV Physical, Collection, Image digital, Image Work, Module / Qualification, Relationships, Text Descriptive, Text Digital, Text Physical, Web. These 12 profiles have been designed to cover all types of materials that may be deposited in OUDA and where the ‘Descriptive’ and ‘Work’ profiles can be broadly related to works as defined by FRBR Group 1 Entities.

Metadata for OUDA materials potentially comes from different sources. The OU/BBC legacy catalogue is a potential source for Audio Visual materials and has been used for a recent project making these assets more visible to aid the re-use agenda in the institution. Metadata is also held in excel spreadsheets and word documents. Metadata is, of course, also held in the library catalogue.

Metadata in the Library Catalogue

Metadata for over 30,000 items could come from the Library catalogue. However, this is not without problems. Firstly, not all items that will be archived in OUDA have a catalogue record i.e. individual units were not always catalogued. Secondly, different items have been catalogued differently i.e. text & A/V items have been catalogued differently as ownership of the cataloguing workflows were split between different teams. Thirdly, cataloguing practices changed over time: fields were deprecated and systematic retrospective cataloguing not undertaken. Finally, and what is of particular interest here, the MARC data needs to be mapped to the OUDA schemas as outlined above.

The first challenge is to identify what MARC records might be usable in OUDA. The catalogue contains over 200,000 bibliographic records. 30,000 of these may be used in OUDA. These records are part of three broad collections.

1. Whilst some MARC vocabularies have been used in the profiles (e.g. MARC Organization Codes) the idea of using AACR2/RDA populated MARC records for OUDA was not appropriate given the array of physical, electronic and conceptual entities being archived.
**OU Module materials** – these are OU teaching materials from 1970 onwards. These materials have changed over time: audio materials have changed from transmitted radio programmes, to audio cassettes and then to Compact Discs (sent in the mail) and subsequently as mp3 files on the VLE. Programmes originally broadcast on television were then sent to students on VHS tapes, then onto DVD and are now streamed on the VLE. Text materials, originally printed and bound are now delivered as pages on a VLE or PDF downloads. Physical kit e.g. home experiment kits and planispheres which were once sent to students often now will have online virtual versions.

**OU Broadcast Programmes** – these are mainstream educational television and radio programmes broadcast peak time e.g. Coast, Timewatch, History of Wales, Thinking Allowed. Historically, these programmes were co-produced with the BBC under the BBC OU partnership agreement. Programmes are now made with other partners too.

**Historical OU Television and Radio** – these are Radio and TV programmes produced for OU students that were not module specific. Also included in this collection are internal institutional recordings e.g. recordings of awards ceremonies & lectures.

These records comprise roughly 15% of the total MARC records in the library catalogue. Separating these items into 3 discrete groups from the rest of the catalogue relied on the consistent application of Open University as a corporate author in the MARC 110 field and the application of Module Codes to MARC records, albeit in different MARC fields. Having experience of navigating, creating and editing the MARC records was also essential in identifying which records belonged to each collection. Once these collections had been identified, data could be added to, and transposed in, the MARC records.

Using MARCEdit to bulk edit MARC records

The actions made in MARCEdit to the records were as follows:

1. Addition of metadata provenance to all records i.e.
   
   040 $aUkMkOU

2. Addition of collection name i.e. Open University Study Materials (24K records), Broadcast Programmes (3K) or Historical OU Television and Radio (4K) e.g.

   490 $aOpen University Study Materials
   
   830 $aOU Study Materials
3  Addition of language code (eng) to 008 characters 35-37. For language module materials the language being taught was added in 041 Language code field e.g.

041  $a fre

4  Addition of module code in discrete field 099 from a different MARC field, e.g.

099  $aA100

5  Addition of year of Date 1 in 008 General description field if not present. Data came from 518 Date/Time and Place Note to 008 field character 7-10.

6  Normalisation of Date of Transmission from 518 Date/Time and Place of an event note to 033 Date/Time and place of an event in accordance with ISO 8501 standard e.g.

518  $a First transmitted on BBC Radio 4 on 29 February 2012 changed to 033 $a 20120229.

7  Normalisation of named contributors in: 100 main entry-personal name, 245 subfield c field, 508 Creation/Production Credits note field, 511 participation or Performer note. Transfer to 700 fields with MARC relator codes and terms e.g.

508  $aProduced and directed by James Hayes
511  $aNarrated by Richard Hammond; contributors Ray Tollefson, Simon Trew, Jack Burke, Joseph Balkoski, Frank Kennard, Adrian Lewis, Steve Zaloga, James Eikner, Gary Sterne.

to

700  $a Balkoski, Joseph. ‡4 ctb ‡e contributor
700  $aBurke, Jack. ‡4 ctb ‡e contributor
700  $aEikner, James. ‡4 ctb ‡e contributor
700  $aHammond, Richard. ‡4 nrt ‡e narrator
700  $aHayes, James. ‡4 drt ‡e director
700  $aHayes, James. ‡4 pro ‡e producer
700  $aKennard, Frank. ‡4 ctb ‡e contributor
700  $aLewis, Adrian. ‡4 ctb ‡e contributor
700  $aTollefson, Ray. ‡4 ctb ‡e contributor
Disaggregation of Publisher fields where 2 had been entered as co-producers e.g.

260 $b BBC/OU

changed to

260 $b BBC $b Open University

Normalisation of duration to ISO 8601 standard e.g.

300 $a Audio-CD $a 30min.,00sec

changed to

306 $a 003000

Edition note. For audio and video material new editions of items were marked in the 084 & 110 field by a Year. This data was duplicated in the 250 field and removed from the 084 field e.g.

250 $a A100/15(1972)

Importing and exporting MARC data from Voyager

Exporting and importing data into Ex-Libris Voyager Library Management System was not unproblematic. Importing 30,000 records into Voyager can take 10 hours, this may have caused other Voyager modules to go-slow. Crashes on re-indexing were common and MARC records that validate OK in Marc Edit may throw errors on re-import.

Identifying records in a collection reliant on fields are indexed in Voyager i.e. 830 field is indexed and 490 field isn’t – which meant duplicating some data in MARC records purely to aid import and export from Voyager.

Any changes in the cataloguing module made to records that have been exported will be over-written on import. This required close liaison with the cataloguing team & system administrator on appropriate times to do the bulk edits.

Bulk edits generally take 2 days to do – 3 bulk edits were done over a 6 month period. Occasional updates of 100 or so items were also undertaken to mop up omissions from the original collections.

Embedding changes into current cataloguing practices

Embedding changes to existing practices went smoothly largely due to the local practice of copy cataloguing from existing records already in the LMS. For instance a new piece of module materials is frequently a
replacement of an old item. The old item may be used as a ‘template’ for the new record. Changes to MARC records were effectively copied across rather than templates created or adding data on scratch records.

Team meetings and special meetings were used to discuss changes and potential impact. Close liaison was required during bulk upload weeks.

Reflections and lessons learnt

Whilst the purpose of the exercise was to move existing data to new more granular fields inevitably some degree of data cleanup was required.

Keep good communications with all people involved!

Learning regular expressions was good professional development and there is plenty of good guidance on using MARCEdit on the web.

Display of some changes in the OPAC (e.g. non-display of MARC relator terms) were not addressed as we were in the middle of procurement of a new LMS.

What’s the impact?

The impact of the changes is yet to be identified. When the metadata is rolled out to the Fedora repository and beyond via linked data, connections between data should be made that had previously not been possible e.g. connections between research outputs and teaching outputs of OU staff.
I work as a cataloguer at Ingram Coutts Information Services, where we supply books to academic libraries around the world, and have a shelf-ready service that provides MARC records for those books. So my thoughts are those of someone who works at the coalface, as it were, mining books or CDs or DVDs or what-have-you for information.

I want to look at some of the shortcomings of doing RDA in MARC, some of the potential gains of doing RDA as Linked Data, and how we might get from one to the other. Linked Data, in a quick definition, is a way of making single units of data available in such a way that computers can read them and can link the data from different sources.

The nuggets I mine are of the type, the film "Rosencrantz and Guildenstern Are Dead was directed by Tom Stoppard", etc. And even as I’m mining them I’m also fashioning them into jewellery, in the form of a MARC record. MARC is so much a part of our grammar, that we perhaps don’t even notice this hidden task of fashioning data, which is wasteful because repetitive. If that fashioning could be more automated, cataloguers would be freed to record more and more impactful metadata – to mine deeper than we currently do. Which I think we could do in a Linked Data environment. You can see another drawback to MARC there – I’ve got a
couple of fields there standing in for the record, there’s useful information in there, but without the scaffolding of all the other bits of the MARC record they can’t be communicated and don’t mean anything – they have zero impact.

=245 10$a Rosencrantz and Guildenstern are dead /$cdirected by Tom Stoppard.
=300 \$a 2 DVDs (approximately 118 min.) :$bsound, colour
=336 \$a two-dimensional moving image$btd$2rdacarrier
=336 \$a video disc$bvd$2rdacarrier
=336 \$a laser optical NTSC$rda
=346 \$a video file$2rdamedia
=511 0$a Cast: Gary Oldman, Tim Roth, Richard Dreyfuss.
=520 \$a "Rosencrantz & Guildenstern Are Dead is a 1990 comedy-drama film written and directed by Tom Stoppard based ... The film was shot in various locations around Yugoslavia. This was Stoppard's debut as a film director, and to date it remains his only film directorial credit"--Freebase.com.
=586 \$a Golden Lion, 1990
=700 1$astoppard, Tom,$efilm director.
=700 1$amotion picture adaptation of (work):$astoppard, Tom.$t Rosencrantz and Guildenstern are dead.
=700 1$aoldman, Gary,$eactor.

Figure 2

Edited highlights of an RDA MARC record (Figure 2). Something old, something new and something borrowed.
Borrowed (Figure 3): the summarization of the content. I’ve cut and pasted that from a website, which is inefficient.
Figure 4

Something old (Figure 4): We’ve got an award note here – that has impact if a person cares to read it, but what would have still more impact would be if you could use it to automatically build a class – “films that have won the Golden Lion” but you’re going to need programmers to jump through all sorts of hoops to do that because my laconic note “Golden Lion, 1990” isn’t obviously the same thing as a slightly different textual string, “1992 Leone d’Oro”, say.
Figure 5

Something’s new (Figure 5): we’ve got these things here little bits of data: content, media, carrier type.
And these (Figure 6): broadcast standard, encoding format and similar. The advantage to us at Ingram Coutts is that we have a range of customers, some want information like that in a note field, some want it in the 300$a, some don’t want it at all. If we input it as granular data we can consistently output it however we want, so the exact opposite of the awards note. This is in the spirit of the RDA, one of the intentions all along was to have “a clear line of separation between the recording of data and the presentation of data”1.
A way of presenting information that’s held as Linked Data is as a graph (Figure 7). For now, the relationships that exist between these things isn’t held in the graph. You can see that the film is linked to Seamus, but not that it’s a song on the soundtrack, a contained Work.
I won't do the whole graph but I'll do a bit of it. The relationship between the film and its director is one we can express in RDA. So I'll put that relationship in and give meaning and direction to that part of the graph (Figure 8).

I've taken that relationship from the RDA element set, which has been approved and is maintained by the Joint Steering Committee and can be found at http://www.rdaregistry.info.

Figure 9

One way in which this is a powerful model is that you have inverse relationships, so the arrow going one way with “has film director”, can give you an arrow going the other way with “is film director of” (Figure 9). The relationship between Tom Stoppard and David Gilmour isn’t a traditional library one. But Linked Data doesn’t care that you’re mixing and matching RDA with some other formal way of representing knowledge. A way to represent relationships between people is known as Friend of a Friend, and that’s what I’ve used here to beef up my library-specific graph. Someone can then come in and add other entities and the relationships between them, a play by Tom Stoppard, the school he went to, adding value to the graph without disturbing the ecosystem, someone can add one piece of data without having to download the entire graph make a change and upload the new graph entire to overlay the old one, as happens now in a record format.

The web knows all these things on our original graph and tons more. What it doesn't know are things in MARC because the web doesn't know how to read MARC. If your users are on the web and your information isn't, your information is hidden from your users, your metadata isn't in their workflow and potentially your library and
its holdings are going to fall out of their workflow. At best, the library remains in the user’s workflow but is a separate domain unlike anything else in their information-seeking behaviour, in the way that the bus is in my workflow on the way to work – I notice the bus because the bus is late, the bus breaks down; the pavement is equally part of my workflow but I don’t notice the pavement because it works, it would never occur to me to take any other route to the bus stop, to walk across my neighbours’ hedges, say. The library’s aim should be to be in the workflow like the pavement, not to be noticed like the bus. Why I think you would want to release metadata in this way – metadata doesn’t have much impact if no one can see it; it’s also going to have more impact if it uses the web as a platform, because that’s where information seekers are, and that’s where people building applications are.

So here I’ve circled the parts of that graph that I would expect to be represented in a library catalogue. There’s a meaningful relationship between the 2 resources represented by the circles – the film Rosencrantz and Guildenstern are Dead and the film Butley – they’re both directed by people who are Fellows of the Royal Society of Literature. You can see it there by following the highlighted nodes.
But making use of relationships like that requires data that isn’t in bibliographic records and requires recursive searches across a highly interconnected domain, which you can’t do in MARC. So I think we need to make sure that the masses of really good data that libraries hold isn’t behind fences, unable to capitalise on relationships like that; isn’t just sitting in silos depreciating in value.

Are there any workflow benefits to a Linked Data environment? The current melange of ISBD, MARC and RDA that we’ve made for ourselves is hard. We’re all very good at it so we don’t recognise how difficult it is but does anybody not have a backlog? Does anybody find it easy to recruit and train staff? Does anybody have more than a tiny fraction of the intellectual works that their library contains discoverable – not the books, but the intellectual works that compose those books. Unless you can answer yes then the current environment isn’t working for you, and if it isn’t working for you it isn’t working for your patrons.

Why would I want to allow others to chain their information on to mine? Well, a frisson’s going to go round the room if I talk about user-generated metadata, somebody in their attic mis-keying saying that War and Peace has the author The French Invasion of Russia. There are certain rules you can sit on top of the framework that makes all this happen that will stop complete gibberish like that going anywhere, and you can put statistical rules in place so that if 9 people are saying War and Peace has author Leo Tolstoy the 1 person who’s saying it has author Barbara Cartland won’t be heard, and you can see where data is coming from – the Library of
Congress, I’ll trust them, a random IP, no thanks, I’ll filter that out. As the final hurdle the cataloguer. The guardian of metadata, putting a tick against the things they like and getting ornery when they find something they don’t, which is the part of the job we all like anyway.

All this happens in the shadow of BIBFRAME, which aims to be, “a format that will ‘accommodate and distinguish expert-, automated-, and self-generated metadata’”.

Instead of beautifully constructed but sterile MARC records we end up with units of data fit to express all the relationships libraries might concern themselves with — between author and book, between cost of journals and the frequency with which they’re cited, etc. It’s the difference between a series of ivory towers with a few cursory rope ladders strung across them, and a big pool of, say, paperclips, paperclips representing units of data, over which you could hover a kind of magnet and pull out a chain of them, the magnet here representing a query in a formal query language.

Figure 12

Now in that puddle of data would be all the information to satisfy an end user’s requirements, whether they actually were Find, Identify, Select and Obtain or something else. And in that puddle would be all the metadata to construct, on the fly, a ‘record’ customised to a user’s needs — you could have an iterative relationship between programme and user such that amidst all the information the web has available on
Orlando by Virginia Woolf, the user pulls out something along these lines (Figure 12) – we’ve got values for the RDA relationships “has film adaptation”, “has author”, and so on. There’s a principle here: Include and postpone\(^3\) – you have to be very clever indeed to predict what metadata is most impactful now, never mind what will be most impactful or desired in 10 years, or 50 years (as that’s roughly the current age of MARC it’s not entirely unreasonable to have that kind of timescale in one’s sights), better to have access to lots of it and postpone a non-binding organisation until it’s needed.

So what now?

With regard to all that metadata that libraries hold one option is to find out how to chisel as much of it out of MARC as possible (Figure 13).

---

Even while it’s in MARC it might be possible to, for example, attach a machine-actionable identifiers to string values. Here the aim is to link your Tom Stoppard to the Web. So I add on an identifier that VIAF knows – VIAF being a name authority service of OCLC. VIAF functions as a hub, so if you talk to it you can talk to, for example, BNB and the National Library of Hungary, while the National Library of Hungary can talk to Europeana. The overall goal being to drop yourself into the Web of data.

Another thing to think about might be – OK, I can add value to my data, I can make it link to kinds of information that aren’t in library records, I can make that value-added data easier to find, but what can I do with regard to
holdings and local information and my local constituency – say my users are interested in Rosencrantz and Guildenstern are Dead, the film, the play – then they’re interested in any material pertaining to that that might be in a library, a gallery, an archive, they’re also interested in which courses have that on their syllabus – at your university and on MOOCs – they’re interested in whether a theatre locally has that in their repertory. And the people at those galleries, those theatres if they have an archive department, are interested in the kind of information libraries have. So it makes sense to think about moving on from presenting that information in a way that only libraries understand, and to do that not simply by proselytising our standards – which seems to me to have very little prospect of success – but by looking at the potential re-use of the data recorded in those standards.

References


Images

goldwhy.com, awomenscentral.com, commons.wikimedia.org
This year four papers for consideration by the Joint Steering Committee for Development of RDA (JSC) discuss possible changes to the Production, Publication, Distribution and Manufacture Statements. This paper presents a discussion of some of the underlying issues and the possible solutions.

The Production, Publication, Distribution and Manufacture Statements (PPDMS), elements 2.7-2.10 within RDA Chapter 2 “Identifying Manifestations and Items”, are used to record the where, who and when related to the manifestation of a work. These statements are endlessly diverse and here is just a small sample of that variety.

In 1841, Antonio Panizzi said all that he needed to say about imprint in instruction Number 27 of his 91 rules.

“Then [record] the place where the book was printed; and in particular cases, as in the instance of early or very eminent typographers, the printer's name to be specified. Next the date : when no date or place is specified, then either or both to be given, if known to, or conjectured by, the librarian ; but in these instances to be included in brackets.”

Panizzi was writing while publishing was still intimately bound with the manufacture of the book, so his concern is with the printer and with date and place of printing. This information is intended to assist the user with identification of the book and may enable a copy of the book to be obtained. It is also significant information for anyone studying bibliometrics or the history of printing.
Panizzi’s specification was elaborated over time into the Publication Statement, reflecting the sequence of Panizzi’s instructions: “...the place the book was printed”; “…the printer’s name”; “next the date.” Subsequent cataloguing instructions also reflect the growing complexity and specialisation of the publishing process, by distinguishing between publishers, distributors, manufacturers and producers of unpublished resources. The consequence is that Panizzi’s Rule XXVII has spawned 186 progeny in RDA, occupying 54 pages.

AACR2 acknowledged functional distinctions in the production process, but RDA separates the functions into clearly defined elements: Production Statement, Publication Statement, Distribution Statement, Manufacture Statement to support more specific enquiries. In principle, RDA makes it possible to isolate information about any one of these functions; for example, to limit an enquiry to manufacture. RDA also makes a clear distinction between unpublished resources (2.7 Production Statement) and published resources.

The other significant change from AACR2, is RDA’s extension of transcription and the removal of exceptions to the rules in AACR2 that had accrued over time. Transcription represents the resource more faithfully, so that a user should have confidence that the metadata provides a reliable surrogate for the resource.

RDA could make these changes because (for the first time) cataloguing instructions were not constrained by requirements of space management. RDA does not expect the information recorded to fit on a 5x3 card, so it does not need to instruct cataloguers to use the shortest possible form of a name or to omit information.

The combination of these changes and the core requirements create a number of issues, which I will examine in turn.

- “cascading vortex of horror”
- the distinctions between published and unpublished resources
- the use and limitations of transcription in relation to user tasks
- complexity and efficiency

Cascading vortex of horror

The cascading vortex of Horror has been summarised as follows, “a chain of instructions the cataloguer may be required to follow in order to satisfy the core requirements when a piece of information is not available.”

The core requirement is to record (at least) the first instance of: the Place of Publication, Date of Publication
and Publisher Name. If the core requirements cannot be satisfied, the equivalent attributes in the distribution and manufacture statements become core.

Here’s a relatively simple example.

**Publication statement when all components are known.**

Publication Statement Place not identified : Canongate Books , 2013

**Statements required when core element cannot be identified**

Publication Statement [Place of publication not identified]: Canongate Books, 2013

Distribution Statement [Place of distribution not identified]: [distributor not identified], [2014]

Manufacture Statement: Kirkcaldy : Danskin Printing , 2013

In the first example, all the core requirements are present on the resource, so the cataloguer can complete a simple Publication Statement. In the second example, the place of publication is not known and cannot be identified; the cataloguer is therefore obliged to create a Distribution Statement. In this case, although the date of distribution is known, the place of distribution and name of the distributor are not; therefore a manufacture statement is required to satisfy the core requirements.

This is a rather laborious process and there is a strong temptation to short-circuit it where possible; for example, by supplying probable information about the place of publication or the date of publication. The consequence is to erode the value of information recorded in the statements. Neither a computer nor a catalogue user can easily distinguish between a genuine place of publication and a place of manufacture supplied as a place of publication. So we have a publication statement in RDA that may be as muddy as an AACR2 statement.

We should also question the utility of recording information in this way, particularly for trade publications. The user does not need it to obtain the resource and is generally not interested in whether the cataloguer has found information or not. The reason for recording “Place of publication not identified”, etc. is to assure other cataloguers that the person who created the record did so with due diligence, so that they do not need to duplicate the work. How much time this actually saves can be questioned.
Distinction between published and unpublished resources

ALA has questioned (6JSC/ALA/Discussion/4) whether it is legitimate to take a one size fits all approach to published and unpublished resources. Published resources, such as books, serials, scores, commercial recordings are (in general) self-describing, whereas, unpublished resources, such as archives, manuscripts, artworks and found objects are generally not. This means that in many cases there is nothing to transcribe, or that the information which is available for transcription may not be authoritative, legible or easy to identify. ALA recommends recording (NOT transcribing) information related to unpublished resources because the information is usually taken from sources outside the resource or it is supplied by the cataloguer.

Uses and limitation of transcription in relation to user tasks

Early printed books, sound recordings, DVDs and many contemporary e-publications may not have conventional title pages and don’t lay out the information neatly for cataloguers. Imprint statements may be presented as sentences. It is not possible to both transcribe this information and separate it into functional statements and sub-elements. Parsing the information into metadata must reduce its value for identification of the manifestation, an important consideration for early printed books in particular. The current solution is to add a note to record the information as it appears on the resource; effectively recording it twice.

An obvious consequence of transcribing data as it appears on each resource is a loss of consistency, which may limit recall. On one resource the printer’s name may be given as T. Este on another as Thomas Easte or even T.E. This won’t help anyone trying to find all resources printed by Mr East. Fortunately Thomas Easte rejoices in a NACO record which collocates the variant forms of his name, however creating authority records for publishers is the exception and is generally not done for trade publishers and modern imprints, so it can be much harder to answer a question such as what did Collins publish before 1970?

There is no provision in RDA to make an authorised access point for Place of Production, Publication, Distribution or Manufacture, or to record variant forms. Nor is there any specific provision to record a controlled form of date of manifestation, so it would not be possible to search for all books published in a particular place.
or all records released in a specific year or range of years.

This falls short of the IFLA cataloguing principles which specifically state that

“The catalogue should be an effective and efficient instrument that enables a user:

4.1.1. to **find** a single resource

4.1.2 to **find** sets of resources representing: …

**all resources defined by other criteria** (language, **place of publication, publication date**, content type, carrier type, etc.), usually as a secondary limiting of a search result.

The diagram illustrates that from the user perspective, the transliterated data mandated by RDA can help to find, identify and select a manifestation or obtain an item when we know the place, name or data associated with production, publication, distribution or manufacture. However, if the information we have does not correspond with what is on the item, it is much less effective. RDA does not allow us to find all the resources published during a specific span of time or in a specific place because variant forms are not recorded. RDA does allow the user to find everything associated with a specific publisher, distributor or agent, if local policy requires these to be established as agents, for example in an authority record.

**Complexity and efficiency**

Finally the RDA instructions at 2.7-2.10 are lengthy and relatively complex. There are 54 pages of text and more than186 instructions. There are inconsistencies to be mastered: places and names must be
transcribed, but dates are recorded; place and dates may be supplied, but names cannot. There are also conditional core requirements and optional additions and omissions and exceptions and alternatives.

Most of this complexity exists in order to split the statement into functions and sub-elements, to avoid creating authorised access points for publishers or places or controlled forms of dates.

**What is to be done?**

Recommendations to address these problems are set out in 6JSC/BL Rep/1 and to a lesser extent 6JSC/ALA/28.

First and most importantly RDA should make a clearer distinction between data that is intended to support the identification of a resource and data to enable resources to be found. It is clear that transcribed data cannot support both tasks. This is particularly true when the purpose of a search is to FIND all resources with a shared attribute.

**More consistent transliteration**

It is recommended that the dates of publication, distribution and manufacture should be transcribed, as is already the case for place and name. This would mean that dates in Roman numerals or presented as words would be recorded as they appear on the resource.

It is also recommended that the division of the statements into separate sub-elements for place, name and date should be discontinued.

**Distinguish between self-describing and non-self-describing resources**

If resources are not self-describing provision may be made for recording information from other sources instead of transcribing information.

Making a distinction between published and unpublished resources may be too crude, as some unpublished resources may describe themselves accurately. It would be helpful to have available a metadata property that makes it clear whether metadata in a given element has been transcribed, recorded or transliterated.
New entities

The two scenarios described below, have some common components, but scenario A has a lighter touch and would be somewhat easier to accommodate in RDA; Scenario B has a much bigger impact on RDA, but offers greater flexibility and more consistency.

Both scenarios assume that the publication, distribution and manufacture statements will be transcribed and that the subordinate elements, Place, Name, Date will be deprecated. This means that there would be no separation into Place of Publication, Name of Publisher and Date of Publication etc.

**Scenario A**

In Scenario A, two new elements are proposed: Date of Manifestation and Place of Manifestation. The purpose of these new elements is to enable controlled terms to be recorded to complement the transcribed data. Controlled terms will enhance collocation, enabling questions, such as, “How many books were published in Edinburgh in the 19th Century?” to be answered.

Place of Manifestation and Date of Manifestation are controlled elements that would enable searching by date or place but variant place names and other attributes of Place or Date could not be recorded.

**Scenario B**

Scenario B recommends replacing all of the PPDM statements with a single element, provisionally named Issuance Statement. This would take transcription to its logical conclusion, enabling complex statements to be reproduced in full. Replacing four statements with a single statement also simplifies RDA. Alternatively, limit the scope of Issuance Statement to publication, distribution and manufacture and retain the Production Statement for resources that are not self-describing.

As illustrated below, Scenario B also recommends the introduction of unconstrained entities for Place and Timespan. Representing Place and Timespan as entities would make it possible to express relationships between Places or Dates and Manifestations, or indeed any other entity. This is a more consistent and extensible approach than defining the Date of Manifestation and Place of Manifestation elements proposed in Scenario A. Entities can have attributes of their own; for example, the calendar or notation associated with a Timespan could be specified and variant forms can be recorded to enable comprehensive recall.
Impact

At the time of writing JSC had yet to discuss any of the changes described in this paper. If accepted the proposals could be far reaching. The introduction of new entities for Place and Timespan would represent significant changes to RDA and are dependent on changes to the FRBR model.

The deprecation of sub-elements for place, name and date in 2.7-2.10 and the deprecation of the PPDM Statements themselves would be major changes to RDA. More authority work would be needed, at least in the short to medium term, to establish authority records for trade publishers, printers, etc. However many corporate bodies that act as non-trade publishers are already routinely established in NACO.
There would be further change to MARC 21 to accommodate the Issuance Statement and new entities.

Legacy data would not be affected, at least in the short term. Retention of the sub-divisions in the legacy PPDMS would potentially enable retrospective alignment with authorised access points for Places, Names and Timespans.

Taking a more consistent approach to transcription is expected to simplify the instructions. It is calculated that adoption of Scenario B would result in a reduction of the number of pages required to explain Production, Publication, Distribution and Manufacture from 54 to 7 and the instructions from 186 to 30, although additional instructions would be required in relation to the Place and Timespan entities. Scenario A would make much less impact on the number of instructions, but the page count would be reduced to around 30. If a distinction is made between published and unpublished resources, the figures for either scenario would be a little higher.

The benefit will be derived from the introduction of a new entity for Timespan and the release of Place from its subject constraints. The definition of these new entities will extend the relationships that RDA can support beyond what is required for recording PPDM. It will create a more flexible and adaptable model which can support the full range of user tasks, including finding sets of resources. It will also make RDA easier to align with event centred models, such as the CIDOC-CRM model used by the museum community.

JSC Papers referred to in this document:

- 6JSC/RDA/28 Creating a priority order for sources of information in Date of Manufacture element (RDA 2.10.6.2)
- 6JSC/ALA/29 Clarifying core element status for “not identified” elements in the Distribution and Manufacture Statements (RDA 2.9 and 2.10)
- 6JSC/ALA/Discussion/4 Transcription issues associated with the Production Statement (RDA 2.7)
- 6JSC/BL rep/1 Simplification of RDA 2.7-2.10

Other documents

**CIDOC-CRM**: *CIDOC Conceptual Reference Model*

**IFLA Statement of International Cataloguing Principles, 2009**
Natasha Aburrow-Jones, the SUNCAT bibliographic Project Officer, gave a paper on the impact of the varying quality and different standards used in cataloging data upon SUNCAT, the Serials Union Catalogue for the UK. She entitled her paper “Deep Impact: Metadata & SUNCAT”, but promised there would be no Morgan Freeman impressions. She introduced her lightning talk with a tweet that she’d seen: “I have a real issue with [data] quality. It’s like beauty – in the eye of the beholder.” Natasha wanted to show this through the experiences of the SUNCAT biblio team.

There followed a brief history of SUNCAT, the Serials Union Catalogue for the research libraries in the UK, offered by EDINA and funded by Jisc. SUNCAT started in 2003; there are now one hundred Contributing Libraries, made up of national, university and specialist libraries, as well as the CONSER database, the ISSN Register and the Directory of Open Access Journals. Libraries range in size from having collections of hundreds of thousands of serials records, to a few dozen. They range from the National Libraries (such as the British Library, the National Library of Scotland), through University libraries (such as the collections held in Oxford, Cambridge, and so forth), to some very specialist libraries, such as the Wiener Library for the Study of the Holocaust and Genocide, the Royal Veterinary College, and the NERC libraries (the Natural Environment Research Council). The SUNCAT Contributing Libraries have a wide geographic range, from Inverness in the north of Scotland, to Cornwall in the South West of England, and even a couple of libraries in Antarctica and South Georgia.

SUNCAT is a physical union catalogue, in that libraries send their data to EDINA, and it goes through a customised normalisation process before being loaded into the SUNCAT database. Libraries send their data in different ways: many send files in the MARC21 format, but data is also sent in UKMARC, MARC21 with UKMARC elements, and non-MARC carriers. Files are also received in Word documents, Excel spreadsheets and Access databases. The non-MARC files have a cross-walk designed for them, putting the data into MARC21 so that it can be loaded into SUNCAT.

Having covered the variety of data carriers, Natasha discussed the content standards used in SUNCAT data. Files are normally hybrid, in that they follow more than one standard. The majority of records are in AACR2, with a growing proportion now in RDA. Changes have been made to the SUNCAT display to accommodate the new MARC tags that are occurring with incoming RDA records. There are also a selection of hybrid records, those coded as RDA with AACR2 elements, and vice versa, as well as records catalogued according to local rules, tempered by AACR2, RDA and even AACR. This variety of data has to be normalised, treating the files as a whole rather than re-cataloguing on a record-by-record basis, so that records can be matched and merged.
Each Contributing Library has a tailored data specification to make sure that the incoming data is normalised so that it can fit the matching algorithm as much as possible, and ensure a coherent display. Some of this is standard across all the libraries (such as changing 6XX$xPeriodicals to $vPeriodicals, only when it is the last subfield in the tag), but most of it is specific to each library. This mainly concerns holdings, as the export file from each system can vary where locations and summary holdings statements are kept, and often vary within each library. Even identifying where the locations are can be a task; they can be in 362 tags, notes tags, local notes, as well as the more standard 852, 866, 856 tags, and these can all be within the same data file from a single library. This may be due to local practices, legacy data issues from previous Library Management Systems, outsourced data which may not follow the same rules as in-house records, and so forth. Records are rejected if they are not a serial, or do not have a local control number, any holdings or a title after the normalisation process; the library is notified of these, so that they can make corrections and improve the standard in their own catalogue, if they so desire. SUNCAT accepts all records from a library, but they do have to contain the basic information that allows for identification of a journal. Records are not viewed through a usual cataloguing interface, but in their raw form; this included showing a record in MARC Communications Format, a sight that many cataloguers do not see.

Records might follow the carrier and content standards, but there is much variation in those records received from different sources for the same journal. This variation will only increase as more records follow RDA, which includes more cataloguer’s judgement than AACR2. The result is a lack of consistency, which will have an impact on record sharing, and trying to match records together. Another impact of non-standardised data is the impact it has on discovery; in SUNCAT, it means that records for the same journal title may not match together due to lack of detail, or different details used to describe the same resource. The non-standard data used results in “satellite titles” in the SUNCAT display – records that have not matched into the main set, and do not make for a clear discovery interface. Every year, SUNCAT conducts a survey, asking for improvements. Every year, one of the most desirable requests is to reduce duplication. If all serials records were catalogued according to the same standard, this would not be an issue.

Serials cataloguing has suffered in cataloguing circles; it can be seen as being too hard to do, or not as interesting as antiquarian monographs, or not as good for the departmental statistics (it is much quicker to catalogue a simple monograph than a serial with a complex series of title changes). This has had the effect of more varied cataloguing in the serials arena than in other areas.

SUNCAT uses an algorithm to match records for the same title so that, ideally, there would be one journal title made up of a set of records, with many holdings attached. The algorithm looks at all incoming records and matches them with existing records in the database. It is points based, looking for the presence of particular
fields used in the identification of a journal, such as the title, author, ISSN, and so forth, and matching the data therein. Once a threshold of points has been reached, a match is determined. SUNCAT has a new interface, which has been designed in-house, to improve the ease of searching. A new algorithm is also being devised, based on making a multi-dimensional radius match. According to the developer involved, all the different aspects of a match are plotted in a multi-dimensional space. A sphere is then made around an item, and matching points in the sphere are used to make a match. This should help improve the matching of poor quality of varying standards of data.

SUNCAT also has some additional functionality which has provoked some interest amongst cataloguers. This is the “Suggest a Match” feature, which will help inform and teach the new algorithm, as well as acting as database maintenance functionality. It allows anyone using SUNCAT to suggest that two or more sets might be for the same title, and generates a report to SUNCAT, which can then be inspected and acted upon. This will aid the reduction of duplication in the SUNCAT database. This functionality, along with the data specifications for each library, covers some of the ways in which SUNCAT is trying to improve the impact of varying forms of serials cataloguing.

In conclusion, it is clear from the SUNCAT experience that there is much variation in the quality and standards used in serials cataloguing in the UK. In an ideal world, the standards would be followed with little deviation. However, this is not an ideal world; there are too many local practices, too many old records which will not be upgraded to new standards. SUNCAT will keep on trying to standardise the non-standard, concentrating on the details that make up the serials catalogue record. To refer back to the tweet that was mentioned earlier, the bibliographic records in SUNCAT are a thing of beauty; it may not be pretty or follow the conventional standards, but the data is all functional and aids the discovery of serials in a single service.

Slides can be found here: http://www.slideshare.net/edinadocumentationofficer/cig2014pres-rev

More information – and the service itself – can be found on the SUNCAT website: http://suncat.ac.uk/search

If you want to contact Natasha with any questions:

Email: Natasha.aburrow-jones@ed.ac.uk

Twitter: @TashaAJ
As soon as I arrived at the Metadata: Making An Impact Conference at the University of Kent, I started to realise that the title of my paper was controversial. The first person that I spoke to said during the conversation, “But there are always second chances”. Not wanting to give anything away, and quietly thrilled by the response, I told them to wait for the talk.

It’s best to begin by giving some background information about where I work, as my daily environment is quite different to perhaps the majority of the members of CIG.

Introduction

I work at RB. Ever heard of it?

What about Reckitt Benckiser, or Reckitt & Colman?

Or Strepsils, Nurofen, Finish, Vanish, Dettol, Airwick, Durex?

RB is the multinational health and hygiene products company that owns and manufactures all of these, and many others. It is in the top 25 of the FTSE 100 on the UK stock market, and its annual turnover is over £10bn. Other ‘powerbrands’ you will recognise include Scholl, Gaviscon, Veet, and, of course, Cillit Bang!
a UK based business with two major locations in England, the headquarters in Slough, and the main development site in Hull, which includes the majority of the factories and the largest research and development (R&D) base. The Hull site also includes the R&D Library and Archive, which is where I work alongside three colleagues.

The R&D Library and Archive includes a book library, subscriptions to journals and online services, multiple legacy collections (slowly but surely being digitised), an environmentally-controlled clinical archive, and a repository of copyright-fee paid research papers. Our services also include our LMS, literature searching services, and training on topics like information literacy, copyright, and using our various systems.

The Library is heavily focused on research papers and subscription stock; we have only approximately four thousand actual books (although many of these are thousand pound encyclopaedias and pharmacopoeias, rather than paperbacks), but well over thirty thousand articles and reports stored in our repository. It is the repository I wish to focus on.

Our repository is a ‘dim archive’, so the next question I asked the conference was had anyone heard of a dim archive? (No-one had) (And no, the thing in Doctor Who is a ‘dark archive’)

**Definitions**

**Dim archive**

A ‘dim archive’ is a controlled repository held with permission of the Copyright Licensing Agency (or ‘CLA’ – the UK authority on copyright), that allows a company to hold a single copy of a document, which can then be shared with selected audiences for certain reasons – specifically with Regulatory Authorities for the purpose of Regulatory Submission.

**Regulatory Authority**

A regulatory authority is a government or official body in a country that reviews medical products for safety, tolerability, and toxicity – essentially the body that ensures that a drug is safe for human use. Regulatory authorities include the MHRA in the UK, and the FDA in the USA.

**Regulatory Submission**

The dossier provided by a company like RB, to a regulatory authority, including information that proves / states our case for said safety.
These dossiers are vital for the business as without approval from the country’s regulator, the company is unable to sell its products.

The dossiers obviously contain a lot of different types of information, but one of the many things that it contains as part of the proof we offer is a number of published research studies that have tested the drug or its ingredients in clinical trials. The R&D Library’s main day-to-day function is to purchase these documents with the copyright permissions that allow us to include them in the dim archive.

**How does the repository work?**

We maintain the dim archive in a piece of software that can ‘lock’ a document once it has been uploaded and approved. This ensures that no-one can alter the text of the document. It also places a watermark on each document on creation, so that it can be easily identified as a library paper. Because only the library staff can upload to this area, we can be sure that only documents that have gone through full copyright (and other) scrutiny are captured in this way. The software also automatically tracks anything that happens to the document. It stores information on who creates it, modifies it, approves it, locks it, and then who views it. This means that if anything does happen that raises a question, we can prove what happened, when, and who was responsible.

However, along with the actual paper being locked, all the metadata that we upload at the same time is also locked – and it is locked to such an extent that no-one, not even the Library manager or our IS department, can amend it.

**Advantages**

The advantages of this process are easy to see:

- For publishers
  - Originality and integrity
  - Trust

By locking one copy, we can approach publishers and rights-holders with the assurance that multiple copies will not be created. We can also tell publishers with confidence that even if an article is stored for many years, its originality and integrity will be preserved just as it was when they first provided it to us.
For our users

◊ One trusted copy
◊ Assurance that link cannot change

For our users, we can provide them with similar assurance that the document in the repository has already cleared its copyright checks, so that there is no chance, if they write a dossier based on a paper in the repository, that we will turn around and say “oops – can’t use that one, give it back”. By locking the document in one place in a set folder structure, staff can also be sure that a dossier’s links will not be broken over time. If you have ever tried to build a website and know the chaos caused by dead hyperlinks, you can imagine what recriminations would ensue if a Regulator received a dossier from RB with links that they then could not access.

◊ One place to search for all content

The other huge advantage for our users is that there is only one piece of software and one ‘cabinet’ folder structure within that to search for these documents. Users do not have to search multiple screens and have multiple windows open to find what papers they have access to, and the chance for duplicated works in departmental silos is instantly diminished.

• For us

◊ One place to search / store
◊ Process set in stone
◊ Can identify our documents – controlled watermarks
◊ Don’t have to chase every use – we are now covered

Finally for us, we too only have one place to look and one process to follow. The fact that the copyright agreements we come to also last indefinitely means that the librarians do not have to re-check levels of copyright every time a document is used and we don’t have to spend our time constantly renegotiating.

Disadvantages

• For us

◊ Can’t correct mistakes found in the metadata
◊ Can’t move anything we’ve locked in the wrong place
Embarrassment and loss of faith from users in the Library staff’s expertise

However there are big disadvantages too. Once something is locked it can’t be unlocked, even for a good reason. The whole point of the system is that data can’t be changed, so even if a paper published in 2010 has been uploaded as 2008 by mistake, we simply can’t correct that mistake. Errors we spot then can’t be corrected before others see them, and this leads to embarrassment and frustration at the loss of respect for our data.

• For our users

○ Difficult to search for articles when there are mistakes with a date or other metadata, which can lead to duplication of costs as new copies are purchased

○ Anything that we cannot agree this level of copyright and storage rights for cannot be used in our dossiers

This can lead to problems for our users when they cannot find that 2010 paper, because the search engine sees 2008 and ignores it. This means that multiple members of staff can buy the same paper over and over again because they believe we don’t already have it, even though the library staff know that we do.

Particularly common places for mistakes to occur are when reports by a committee are filed under both the committee’s full name and their acronym, some books sections are filed under the chapter author’s name and some under the name of the book’s main editor, and a journal article lists two dates – date of publication and date of acceptance – and we type the wrong one. Another irritation is when an addendum or correction to a paper is printed in a later edition of the journal and we have one file with the paper and a separate file just for one sentence on a largely blank page.

An example of a favourite mistake

One place where I am aware of an especially irritating mistake is a host of monographs from a book entitled The Handbook of Pharmaceutical Excipients, the most recent edition published in 2012. The lead editor of this book is Raymond C Rowe. Our naming convention for files is lead author/editor surname, first initial and date. This means that these monographs should be filed “rowe r 2012”. There are about twenty monographs from this book in the repository; most are under “rowe r 2012”, but a significant number are instead under “rowe c 2012”. At some point in the past a member of library staff has seen the initial C in Raymond C Rowe, and uploaded these by error. In most library catalogues this would take seconds to change, but in our repository these are stuck under C for the foreseeable future.
So with no ability to correct mistakes in the metadata, the best thing we can do as library staff is to avoid them in the first place.

How to avoid mistakes

I talked to my team before giving this presentation and between us we compiled the following list of tips for how to avoid making mistakes when uploading papers:

1. Take responsibility – don’t rely on the approvers to catch everything
2. Learn where mistakes are likely to happen – so you can pay attention to those bits when you are approving
3. Stick with precedents and conventions – even if you prefer today’s answer
4. Don’t trust deliverymen – the coversheet isn’t always right
5. A wrong and a right don’t always solve the problem – think about the conflicting priorities of correction and duplication
6. Concentrate!

7. Get into a rhythm – but don’t lose your concentration

8. Communicate amongst the team – discussion and consensus

The first two points emphasise the twin roles of uploading and approving. The person who loads the metadata in the first place should take responsibility for being as accurate as possible and not being over reliant on the approver to catch mistakes. If when you upload you get 80% right and trust the approver to catch the other 20%, what happens if they only catch 15%, or even 19%? Uploaders must always strive for 100%; the approver’s job should be to sign off – not to make corrections. However the approver is the last line of defence and should concentrate completely on letting no mistakes slip through. One of the techniques I use when approving is to learn where my various colleagues often make mistakes. One of my colleagues tends to copy and paste the pagination from the citation, rather than check the actual pages, which can sometimes lead to inaccuracies – so when I approve their documents I pay particular attention to that field. This isn’t being disrespectful to that colleague – it’s doing the job of the approver in an intelligent way – they always check the spelling of my keywords as I don’t have a science background and sometimes typos creep into my medical terminology!

Another tip we employ is to decide on a precedent and then stick to it. If I was able to tidy the repository I could spend hours standardising items so that all book sections went under the editor rather than the chapter author, or all organisations under their accepted acronym rather than their full title, but in practice it is always more efficient to match what has come before. If we have papers from the World Health Organization (I know – it pains me to spell it with a ‘z’) or ‘WHO’, which are filed under WHO, then all subsequent papers should also be ‘WHO’, even if we all agree we’d rather use the full name. Why – because then there is a single consistent place to search for those papers. The benefits of good communication and consensus amongst team members here can’t be overstated.

Similarly if you have a paper labelled 2008 when it should be 2010, one solution would be to just upload another version as 2010, but that creates duplication in the repository. It doesn’t necessarily mean that the users will always link their dossier to the more accurate version, and it could even be construed as dishonest when we have given those assurances of ‘one single copy’ to the publisher.

**Summary and conclusions**

Sometimes there are no second chances to get metadata right, and there may be good reasons and advantages for that. When you find yourself in that situation, don’t just hope that people won’t notice the
mistakes, have answers prepared for any questions and complaints that they raise. This can help to mitigate
the embarrassment (or at least prepare you for it) and ensures that your team is united in its approach. Be
helpful, teach searching strategies that will assist your users to pick up mistakes or find what they want despite
them, and use this to develop a reputation for yourselves as a team that people can come to if they have
trouble finding anything. This makes you look like the experts who can navigate the pitfalls of the system, rather
than the fools who can’t use it properly.

Above all, as librarians, if you can’t always be right – always be helpful!
We all know that cataloguers produce great quality metadata but we also know that our users seldom turn to the library catalogue as their first, or even second, source of information. However, many users are active on social media on a regular basis, keeping in touch with their peers and their institutions. Libraries are taking advantage of this by promoting themselves and their collections on various social sites. This article will discuss the use of the website Pinterest to promote the Library Science Collection at Cambridge University Library. Pinterest (http://www.pinterest.com/) is an online pin board where users can share images of interest with their followers and at Cambridge we use this to share our metadata in visual form.

The Library Science Collection at Cambridge (http://ullibrarysciencecollection.wordpress.com/) is a result of the library’s legal deposit status. This dedicated professional collection is made up of approximately 2000 items, mainly monographs and journals but also a growing e-book collection. After taking over the management of the collection one of my key goals was to advertise this valuable professional resource to both Cambridge librarians and the wider information world. The collection now has an established social media presence including a blog, Twitter account and Pinterest site.

**Rise of the visual web**

Pinterest has become increasingly popular in the last few years as part of the phenomenon of the visual web (Chan, 2013). Visual websites are overtaking text based sites as the main method of communication on social media. In recent years sites such as Instagram and Snapchat have launched to huge acclaim and almost instant popularity. Analysts claim that we are moving away from a text based web towards something more visual (Orsini, 2013) and people are designing image based websites as a way to tap into this trend. You need look no further than some of the newer templates on popular sites such as Wordpress to see where this trend is taking hold. Couple this with the general decrease in people’s attention spans and it becomes clear that websites now need something different to draw users in and engage them with the content. In addition to this, image based sites offer a better display on mobile devices. Surveys show that people are accessing increasingly large amounts of information on devices such as smart phones and tablets which may have problems with text based sites (Pew Internet). As information providers we need to be aware of this when designing our web presence.

The visual web has several advantages over its text-based counterpart. Think back to a time when you have done a web search only to be confronted with one page full of densely packed text and another page full of...
images. There you can start to see how difficult text is to absorb quickly whereas carefully chosen images can make an almost instant impact.

**Using Pinterest at Cambridge**

So how can we use this knowledge of the visual web to make our metadata more discoverable?

As mentioned, Pinterest is a collection of online pin boards where users can upload, link to and share images which then link back to content. An example of a Pinterest board can be seen in Fig. 1. Visitors to the site can browse the images and then click through to visit the original content such as a blog post or news item. Libraries can use the same mechanism to share their metadata with users in a visual form.

![Figure 1](image)

Currently at Cambridge we are using Pinterest to create an online new books display. Cover images of the books are displayed on our pin board and users can access the catalogue record by clicking on the image. This has obvious benefits over the traditional new books display still used by traditional libraries. With an online display the physical books are still accessible for use so users are not restricted from accessing the actual material. Users no longer need to be in the library building to see what has been added to the collection, opening it to a wider professional audience. Another advantage is the site’s simplicity for both librarian and user. It is easy to navigate, even for those new to the format, and does not require registration to browse. The images posted display in online image searches, further aiding discoverability.
Constructing a Pinterest board is a simple process. Obviously the first thing you need to do to set up a page is register for an account. Before you do this it is worth finding out if your institution already has an existing account which you can join as this can help to grow an audience for your site. Accounts can have multiple boards on a variety of subjects. Each individual image on these boards is known as a 'pin'. Pins can be uploaded from a computer or taken directly from a website. The easiest way to do this is to install Pinterest's Pin It browser extension (Fig. 2). This extension is available for most browsers from the Pinterest homepage and selects images with one click.

The crucial component of a Pinterest site is to link the image back to the content you wish to share, in the case of Cambridge this would be the metadata of the book represented by its cover image. When you pin an image from the Internet Pinterest automatically links back to the original content. As we use Library Search at Cambridge, Pinterest links to the main catalogue interface as a default. A small amount of editing is required to make sure that cover images link back to the stable url of the item rather than the main homepage of the catalogue. This is achieved by using the Pinterest edit interface. If you select the small pen symbol (Fig. 3) on the pin this takes you to an edit screen where you can make the necessary changes to the source URL.

As the image is uploaded Pinterest will provide a short description in the description box (Fig. 4). In order to increase metadata exposure, at Cambridge we expand on this to include the basic bibliographic details of the book - full title, author and classmark. This provides a quick reference to users without having to click through to the catalogue, although this is still the main aim of the project. When the pin is complete it functions as a visual link to the catalogue.
The Library Science Pinterest board has been extremely popular, gaining nearly four hundred followers in the first few hours after its launch. Since then follower numbers have continued to climb and engagement via likes and shares has been positive. Individual pins and whole boards can also be promoted easily. For a site which requires minimal work to produce this is an excellent outcome which has helped to promote the visibility of both the collection and its metadata.

Copyright

Like many other image sharing sites, Pinterest has been the subject of some copyright discussion. It goes without saying that libraries using Pinterest need to be mindful of copyright, particularly when using cover art as these are of course copyrighted works in their own right.

Prior to starting the Pinterest project at Cambridge we consulted the University Copyright Department who cleared us to use images that were available through our catalogue interface as long as they linked back to the source of the image - the catalogue record. Not all titles in Library Search have cover images so unfortunately we cannot include images for these unless we seek individual permission from the copyright holders. We decided early on in the project that this was not practical and so do not include these books on our board.

If you cannot obtain clearance from your institution’s copyright department or are in any way unsure if you can use the images then the safest policy is to avoid using them.
Other uses for Pinterest

Even if pinning cover images is not an option for your library there are alternative ways of using Pinterest to make your metadata more discoverable online.

Topic and resource boards are easy to assemble and can make a great visual impact. Using images you have either created yourself or found an appropriate Creative Commons license for you can create eye-catching pins which you can then link back to your metadata. Topic boards focus on a theme and can take advantage of popular news stories such as the anniversary of the First World War. Resource boards use a similar approach but centre around popular academic topics such as *How to write an essay*. By using Pinterest in this way you can create a graphic reading list of items and ensure that your metadata is noticed.

Top pinning tips

If you do decide to pursue Pinterest there are some important points to bear in mind. As with other social sites Pinterest allows account holders to write short biographies of themselves or their institution. In addition there is a chance to compose a short description for each board. Write these carefully and try to include relevant keywords to increase discoverability. Choose or create your images with care to create the maximum impact on both Pinterest and the Internet in general. The aim is to make people stop and look in an environment which is increasingly becoming image dependent.

Conclusion

Pinning images with links to our metadata has proved hugely popular at Cambridge. It provides us with a way to tap into the trend for the visual web using metadata - something not traditionally thought of as picturesque. Hopefully this helps us to make more of an impact with our metadata on social media.

Bibliography


Background

The Record Enhancement to Aid Discover-ability Initiative was developed in October 2012, in response to a number of stimuli that began in the late-2000s when at DMU we first had demonstrations of discovery systems, and can definitely be pinpointed as starting in May 2010, when:

- DMU appointed a new Vice Chancellor, following the retirement of the previous incumbent: This VC clearly has the student experience at the heart of his vision
- In September 2010 I attended my first CIG conference since the year 2000: This was immensely inspiring, particularly Gary Steele’s presentation on applying LCSH to bib records based on those that had already been assigned by others
- In late 2010 the High Visibility Cataloguing initiative took off, and I contributed a post on the role of the cataloguer in the 21st century
- In early 2011, along with many other cataloguers, I engaged in a worldwide Twitter conversation, #catbkchat, discussing the ground-breaking book “Conversations with Catalogers”
- The CIG12 conference in September 2012 proved to be another prompt to develop the work of my own cataloguing team

Thus, the READ-ability Initiative was born.

The rationale

The idea of the READ-ability Initiative was to bring together a number of different activities the cataloguers were already doing, and to increase the programme of discoverability. I wanted:

- To show how the Bibliographic Services Team was committed to improving the experience of students and other library users when interacting with the library catalogue and the library stock, recognising that users want information contained within our resources, not the resources per se
• To articulate and understand what Bib Services as a team did and how what Bib Services did helped the users.

• To promote the work of Bib Services to the wider library staff: Since we lived in the basement, this was a crucial part of raising our profile

The READ-ability Initiative focused mainly on the work of the cataloguers, so the PIC Project (Protecting the Integrity of the Catalogue) was developed as a companion to READ-ability. It:

• Focused on pulling together much of the work done by other colleagues in Bib Services

• Recognised that most of the work done by these colleagues helped to protect the integrity of the catalogue, which in turn helped with discoverability of information.

It should be noted that the regular activities of the team, like ordering, cataloguing, classifying, processing (and hotlinking [hyperlinking] where appropriate) new books, streamed videos, DMU theses and new and amended journals titles, already contributed to both the READ-ability Initiative and the PIC Project. So, the focus of the two programmes was on new and different activities, which, wherever possible and appropriate, would become part of the regular work of the team.

The detail

READ-ability identified the following elements for inclusion in the programme:

• **LCSH enhancements**: Up to now, LCSH were not routinely added to records, and were not checked where they already appeared in records. Our OPAC was quite outdated and we were still maintaining its bespoke subject index: implementing a new discovery system would mean that we would lose all this work, and LCSH would be the main subject access point (as long as terms appeared in the bib records!)

• **LC Name Authorities enhancement**: Name headings were routinely checked and added to incoming new bib records, but there were many thousands of older bib records which had not benefited from this in the past: these would now be tackled

• **E-book record separating and upgrading**: E-books had been added to the equivalent hard copy records. With the increase in purchase of e-books, it was decided that these records should be separated
• **Continuous programme of hotlink checking and upgrading**: In our desire to provide users with as much access to information as possible, we had always added hotlinks to our bib records where we identified there was an exact replica of the hard copy freely available on the internet. These links were notoriously fragile, and subject to breaking when, for example, a government department changed name (as noted at the open mic session at CIG10). Checking these links became an integral part of our work.

• **Completing re-classification to DDC23**: This is also an integral part of our work, and with each new edition of DDC we develop a programme of work to re-classify our stock as appropriate. [See our Strategies for Re-Classification, which explains this in more detail]

• **Participating in DORA REF Submissions Project**: Our Institutional Repository Officer welcomed our involvement in the run up to REF2014, helping to input new records to the system so that our submission was ready in time.

• **Contributing to DORA Record Upgrading Project**: Records that had been added to the repository in the early days became part of an upgrading project to ensure authority control and subject access were capitalised upon.

• **Acting upon Typos of the Day for Librarians**: Subscribing to this international initiative provides a focus for correcting spellings in our bib records.

• **A thirst for continually seeking record improvement opportunities!**

The PIC Project included the following elements:

• **Re-classification**: Aside from the intellectual benefit of the re-classification programme, the physical re-labelling of books is essential. Today’s spine labels are more durable, easier to read and we have a robust process for standardisation. This also makes shelving returned/used items easier and quicker. Re-labelling of stock found through the stockcheck process also yields the same benefits.

• **Hotlink verification**: A robust process for checking the accuracy of our hotlinks ensures our links work, and that any problems are fixed quickly. Library staff help us keep them up-to-date by reporting to us, so that we can repair the broken link.

• **Stockchecking**: When completed this will ensure that the catalogue accurately reflects what’s on the shelves and that what’s on the shelves is in the right place. Some examples of problems found included books being found during the stockchecking process where the newest spine labels had fallen off,
OPAC and the book spine being at odds with each other, and the item effectively lost. Books were also found to be mis-shelved, and in some cases RFID tags were incorrectly (or not at all!) programmed. All these issues will be ironed out by the end of the stockcheck, but it is important to carry on the process as a rolling programme of improvement, particularly since major issues will have been resolved during the first stockcheck run, thus making a rolling stockchecking programme a quicker activity.

- **Withdrawing**: We have a quick and efficient process of withdrawing items weeded by subject librarians, and this helps to keep our stock up-to-date and our catalogue accurate.

- **Missing items process**: We have a process in place whereby items that are misplaced on the shelves (for whatever reason) or books that have genuinely gone missing will be identified quickly and appropriate action (e.g. withdrawal, purchase of replacement) taken.

- **Repairs**: Issuing books to the repairs number means they are always accessible. Making quick decisions about whether to repair, withdraw, bind or replace means that appropriate action can be taken quickly and books not out of use for too long.

- **Re-locations**: Our responsiveness to particular decisions allows us to take prompt action to relocate stock. This is vital so that accurate locations/loan periods appear on OPAC and helps users easily locate what they're after. Examples include the changing of short loan items to 7 day or normal loan, journals from the Store moving into the Special Collections area, and the VHS moving into the Stack.

- **Binding**: Our comprehensive programme of binding vulnerable loose copies of journals ensures that these are easily found on the shelves, through all issues for a certain time period being located together and spine information being consistently applied.

**Progress to date (September 2014)**

- **LCSH and Name Authority enhancements**: The numbers of bib records lacking these accurate access points at the time we started the initiative was staggering: We have made great progress on this, but there is still much work to be done.

- **E-book record separating and upgrading**: This is now complete.

- **Re-classification to DDC23**: This programme of work has seen us complete the “quick wins” – basically, re-locations and discontinuations – and some areas that we have identified as being important to our users. However, there is still much work to be done.
• **Typos of the Day for Librarians**: We still participate in this project, and undertake our own random checking at every opportunity.

• **Hotlinking**: We have now linked our resources through to the Wayback Machine, so an annual check is now sufficient, but where this is not possible, library staff still inform us of broken links. This has also meant that cataloguers are no longer involved in the checking for up-to-date links.

• **Stockchecking**: The first stockcheck has been completed for most of the collection, and major problems ironed out. We are awaiting more state-of-the-art kit before starting the whole process in earnest again. In conjunction with the Missing Items Process, and the careful management of stock return by the Shelving Team, this has meant that if it says it’s available on OPAC, then it really is!

• **Withdrawing**: The number of withdrawals seems to have increased enormously in the last year, but our efficient process is working well, and our initiative to “recycle” unwanted stock helped the library to earn a gold award in the [DMU Green Impact Project](#)

• **Missing items process**: This is an on-going process that continues to work well.

• **Repairs**: The number of items going through the evaluation stage has increased somewhat, but the actual number of books being selected for repair is decreasing. The process involves much liaison between the Bib Services Team and the subject librarians, and they now have a more efficient process for checking and evaluating, which in turn makes the whole process quicker.

• **Re-locations**: Another on-going process that continues to work well.

• **Binding**: This process has been subject to a review, and is running efficiently. Working closely with the Archivist has meant that it has been possible to do further work on vulnerable journals in the Special Collections.

**Associated work**

In April 2013, we [opened our basement office doors](#) to the rest of the library staff, and based this visit to see our team in action on READ-ability and PIC. This event was very well received and helped to cement relationships between library staff, and explain the mysteries of the basement to staff.

In February 2014 we finally upgraded our OPAC to the latest version: The Bib Services Team used this opportunity to again invite library staff down to the basement, this time to [exchange experience](#) in the use of...
the new interface. This proved to be an extremely popular event, and a format that worked well.

Way back in 2008 the team worked on “streamlining” our processes: We’ve always been good at reviewing our processes and tweaking them to adapt to changes in the environment (new software, changes in external processes etc.), but in April 2012, we set up a number of more formal process review groups. These groups worked on processes, with the aim of looking at them to see what was actually being done, that it was fit for purpose and that everyone was doing it consistently. If it appeared unfit for purpose then the group would suggest improvements that could be made, and would communicate these changes to everyone involved in doing the task via written documentation and a short training session. The new process would then be adopted, and reviewed at regular intervals.

The missing items, withdrawal and repairs process have each been subject to a process review, and the binding process was one of the more recent processes to be reviewed. As a result of these reviews, processes are now even more efficient and timely.

The current and future landscape

In September 2014 we finally released our discovery system and members of Bib Services, along with other close colleagues, offered training sessions for library staff. The discovery system is where we expect most of the work of READ-ability to have an impact. In conjunction with a new, google-like search interface, we expect users to find the information they need more easily, because of the work we have done on discoverability.

Within the scope of the READ-ability and PIC programmes, we have identified several areas for improvement: In most cases we are waiting for the right time to launch these on the wider library staff. This includes:

- The potential re-classification of our Shakespeare collection
- Several new process reviews (e.g. journals, e-resources) to ensure discoverability
- The potential, consistent addition of LCSH to journal records
- The separating of e-journals from their hard copy bib records
- The consideration of including hyperlinks to e-content (e.g. TOCs, publisher’s websites etc.).

Some of these, I know, will be superseded by new systems and/or software offers that we take advantage of, and I’m also pretty sure that new opportunities will present themselves! We will also have the challenge of
increasing resource budgets, and decreasing staffing resources, all of which make improvements at the same time more important, but also more difficult to implement!

And from each CIG conference I attend, I take away yet more ideas to work on!
Introduction

Founded 1965, the University of Kent is a research led University with 19,275 students and 3,152 staff based at campuses in Canterbury and Medway and centres in Athens, Brussels, Paris, Rome and Tonbridge. Multi-disciplinary in nature, the University has 3 faculties comprising 18 academic schools and 3 academic centres. Top 10 for overall student satisfaction in the National Student Survey 2013 the University was 28th in the 2014 Complete University Guide, 20th in the Guardian University Guide 2014 and 24th in the UK for world-leading (4*) research Research Assessment Exercise (RAE) 2008.

Information Services (IS) is a converged service comprising Customer Support, Library Collections, IT Development and Planning & Administration providing libraries in Canterbury, Medway and Tonbridge. IS is the custodian of a number of special collections, including the British Cartoon Archive, and works in partnership with Canterbury Cathedral Library and Rochester Cathedral Library.

I joined the University as Head of Collection Management in August 2011. Prior to this my background was in academic libraries and professional associations including as Cataloguing Manager at the University of Aberdeen and Manager – Content & Information Architecture at the Royal Institution of Chartered Surveyors (RICS). I have been a member of the CILIP Cataloguing & Indexing Group Committee since 2009 and was elected Chair this year. However I should stress that the views expressed in this article are my own and not necessarily those of the University or CIG.

This article also reflects that I have a particular approach to the challenges facing metadata teams and the wider information profession, although I respect the views of others who may disagree with me. I usually describe myself as a glass half-full person, but in doing so acknowledge that it could be fuller of something nicer. Finally I would say that although I have high aspirations I can’t claim to be living the dream – yet!

The challenge

The challenges facing the Metadata team at the University were, and remain, many and varied. Like all libraries the expectations of our students, staff and wider community of users are increasing. The drivers behind this are complex and ever changing but there was, and continues to be, a clear need to respond to them in order to ensure not only the ongoing value of the library but also the positive profile of IS and ranking of the University.
Our response was focused in four key areas:

- Building additional capacity to cope with growing volumes and diversity of materials requiring description;
- Improving existing services to match user expectations;
- Introducing new services to demonstrate continuing and growing value to the University;
- Maintaining quality to ensure a consistently excellent user experience.

Working with the IS Senior Management team, I was keen to ensure that our approach:

- Leveraged the professional expertise of the Metadata team;
- Built on and developed those existing skills;
- Empowered both the team and individuals within it.

The opportunity

While the challenge was daunting there were significant opportunities. There was, and remains, a clear commitment from the University in its strategic plan to:

- Offer an inspiring student experience;
- Produce innovative world leading research;
- Operate in an effective, efficient and sustainable and professional manner.

IS matched this commitment with its own vision of the University:

> With knowledge exchange and stimulation of creativity at its heart, where our community enjoys convenient and effective access to information and technology, and where we all work, learn and collaborate how and when it suits us, wherever we may be.

Seeking to:

> Provide leadership, expertise and outstanding services, which empower our University community to exploit knowledge, information and technology to achieve excellence in all our research, teaching, learning, enterprise and business activity.
And:

*Be a transparent, accountable and sustainable organisation where professionalism, open-mindedness and pro-activity help us anticipate and respond to the diverse needs of the University community and support and develop our own staff. We are able to achieve our objectives by being inclusive, customer focused, proactive, innovative, flexible and responsible.*

A clear IS Strategy, a challenging but achievable IS Operational Plan, an effective governance framework providing both oversight and support and a rigorous project management methodology set the tone. The IS Senior Management team appetite for (appropriately managed) risk and understanding of the need for change management created a positive culture for development. Finally a brilliant team enabled us to take forward an ambitious and transformational programme.

**The programme**

That programme was a 3 year shelf-ready strategy simplifying, streamlining and improving our internal processes, and developing our strategic relationships with the private sector, for purchasing and providing access to library materials to:

- **Improve service**
  
  *Making newly purchased items available to library users as soon as possible;*

- **Ensure sustainability**
  
  *Providing a flexible model that allows us make newly purchased items available within the same timescales regardless of the number of orders submitted or items delivered;*

- **Enhance data quality**
  
  *Ensuring our bibliographic data supports: enrichment with cover images, previews and tables of content; searching and suggestions within the catalogue;*

- **Deliver cost and efficiency savings**
  
  *Enabling us to redeploy IS staff time currently used to catalogue, classify and process newly purchased items to focus on priority areas which will improve the student experience and support research.*
The objectives

We developed a set of user-centred objectives for the programme, outlining how it would respond to the University and IS strategic plans by:

- Improve the service offered to library users by making newly purchased items available as soon as possible
  *Reducing the time taken to make an item available by 50%*;

- Improve the service offered to library users by RFID tagging all items in the Templeman Main Collection
  *Making it easier and quicker to borrow and return items via the self-issue machines, to locate misplaced items, and to identify missing items which should be replaced*;

- Improve the service offered to library users by increasing the quality of the catalogue data
  *Ensuring all items at the Templeman and Tonbridge Centre Libraries are discoverable through our resource discovery tools and linked by consistent classification, names and subjects
Using shared services as a source of data that can be used to effectively and efficiently link University systems and support further shared services; outsourcing this work to reduce the cost by 98%, ensure that this project is completed, and that it is completed in a shorter timescale than would otherwise be possible*;

- Introduce standard classification for all new Core Text and Main Collection items at the Templeman and Tonbridge Centre Libraries
  *Maximising the cost effectiveness of outsourcing cataloguing, classification and processing of newly purchased items*;

- Introduce shelf-ready services for new items purchased from our principal suppliers at the Templeman and Tonbridge Centre Libraries
  *Realising the benefits of releasing a minimum of 2 FTEs of IS staff time and reducing the cost of making an item available by 50%*;

- Reclassify the Templeman Core Text and Main Collections in line with standard classification where appropriate
  *Outsourcing this work to reduce the cost by a minimum of 33%, ensure that this project was completed, and that it is completed in a shorter timescale than would otherwise be possible*;

- Improve the quality of the catalogue data
  *Using shared services as a source of data that can be used to effectively and efficiently link University systems and support further shared services; outsourcing this work to reduce the cost by 98%, ensure that this project was completed, and in a shorter timescale than would otherwise be possible*
The Projects

We identified four projects that would deliver the objectives of the shelf-ready strategy:

Establish Standard Classification

A series of meetings between key stakeholders within IS agreed the use of standard Library of Congress Classification with existing local practice for cutters.

Introduce Shelf Ready Stock

This brought representatives from the Academic Liaison Services, Acquisitions, Learning & Research Systems Development and Metadata teams together with sales and technical teams from two book suppliers to develop an order process whereby:

Orders are placed on the supplier website, the supplier providing basic bibliographic and order records which are loaded overnight into the library management system for confirmation;

The supplier provides full invoice and bibliographic records which are loaded into the library management system on receipt of the items.

Reclassify and RFID tag the Core Text and Main Collection at the Templeman and Tonbridge Centre Libraries

This saw the Lending Services, Metadata and Stock Control teams partnering with an external contractor to undertake this major 7 month project ensuring that no item was unavailable for more than 24 hours.

Develop Quality Control for Catalogue Data

Through this project the Metadata team developed a new exception-based model for ensuring the accuracy and consistency of the library catalogue working at collection level to avoid the need to check individual items, prioritising and scheduling their interventions to minimise disruption to library users and maximise impact.
The outcomes

The shelf-ready strategy has had a significant impact on the user experience. Urgent items now reach the shelves within 1 working day of receipt, with all other new purchased items being made available within 5 working days of receipt, halving the previously estimated average of 10 working days. Current cataloguing backlogs have been eliminated, from an estimated 5,000 items, while significant progress is being made on providing metadata for previously undiscoverable special collections.

Library users are also benefiting from 681,117 upgraded bibliographic records. Enhancements delivered include the:

- Adoption of Standard Library of Congress Classification facilitating browsing and retrieval;
- Addition of appropriate RDA field;
- Expansion of abbreviations and replacement of Latin terms in line with RDA;
- Introduction of authority control for names and subject headings ensuring consistency and providing appropriate cross references;
- Correction of fixed field data, indicators and subfield codes improving display and retrieval.

The newly established framework for ongoing quality control has provided a matrix of reports and documented workflows to address known issues, scoped and prioritised according to impact on users and achievability.

Library users will see further benefit from the shelf-ready strategy following the launch of LibrarySearch, the University’s discovery layer, in January 2015. Testing by both users and the project team has demonstrated both that our metadata effectively supports the features of the system and that many of the issues with metadata experienced by other libraries have been addressed.

Less visible to library users is the work associated with the reclassification (in itself a measure of the projects success) which saw:

- 407,805 items RFID-tagged;
- 287,993 items reclassified;
- 500,593 items moved;
- 54,235 items previously undiscoverable made discoverable.
The impact

The impact of the shelf-ready strategy led by the Metadata team extends beyond the outcomes I have highlighted to include the improved discoverability of, access to and management of the collections. The scalable and sustainable workflows introduced have allowed the team to provide metadata records for a growing volume of acquisitions. There was a 25% increase in additions to stock in 2011-12 and a further 12% rise in 2012-13.

There has also been a significant impact for the Metadata team. Previously accustomed to a permanent state of crisis management the team has seen a reduction in complaints in the form of requests for unavailable items and is now focused on the delivery of value-added services. In addition to a growing focus on the University’s unique and distinctive collections, the team have active involvement in a broad range of internal and external projects and initiatives including:

- Cataloguing to archival standards, including materials for the British Cartoon Archive;
- Responsibility for Kent Academic Repository (KAR) metadata, including that used to support academic promotions and for the deposit of research theses;
- Creating metadata profiles for the Kent Research Information Management System Online (KRIMSON) including that required for integration with KAR;
- Exploring the metadata requirements associated with research data management and funder reporting.

All of these have seen the Metadata team developing and demonstrating new skills, building on and complementing our core professional competencies.

The future

A key indicator of the impact of the Metadata team is the current IS Operation Plan. This identifies the team as central to 21 (15%) of the projects listed. The team are planning to be involved in a further 11 projects on the IS Library Collections 2014-17 Roadmaps, under the strategic priority resources, for:

- Partnerships, Equality & Diversity

  Exploring partnership and collaboration, procurement, unified resource management and volunteering strategies;
- Learning & Teaching
  *Examining active collection management;*

- Scholarly Support
  *Enabling data exploitation;*

- Special Collections
  *Leading digital asset management, digitisation and unique collection highlighting*
  *Supporting accreditation and publishing;*

Through these and other projects the Metadata team will harness their (and IS’) growing relationships with other IS teams, academic and professional service departments, regional partners, funders and consortia. This will see the team honing their skills to:

- Exploit library resources, management information and research data to deliver business intelligence;

- Play an active and leading role in areas relevant to our expertise.

- Build on our reputation for success and continue to:

- Develop new services for our users;

- Add value for the University and our communities;

And most importantly to continue to make an impact.
Information Services, at the University of Kent, delivers an amalgamated IT and library service to over 19,000 students. The largest campus is located in Canterbury, Kent with further campuses and centres in Paris, Brussels, Rome, Medway and Tonbridge. Within the Information Services structure the Metadata team is part of the wider library service as illustrated here.
In late 2012 we both saw an advertisement from the University looking for a library assistant to join the Metadata team in the Templeman Library. At the time Clair was working for Canterbury Christ Church University as a Senior Library Assistant, and Josie had the same job title within Bexley Libraries. The advertised job description for the role at the University was something we both were unsure we could live up to, but similarly we felt that it was time for a change and were both up for the challenge. Neither of us has traditional training in cataloguing or had very much experience of working in academic libraries (Clair had only been at Canterbury Christ Church University library for around 14 months; having been working for Kent Libraries & Archives before this). Interestingly, we both came from very similar backgrounds, with almost 20 years’ experience working in public libraries between us. We came to interview, and, as you may have guessed, were offered the job!

We joined the University at the beginning of 2013 (in January and February respectively), arriving at a period of considerable change due to staff changes and new working practices; following a round of voluntary severance the Metadata team at the time consisted of just one member of staff. Information Services was just completing a shelf-ready project, and was preparing to start a reclassification project of the entire library.

Our first days were both exciting and overwhelming. The scale of the campus was daunting and arriving to this and a library with four floors and two wings was a distinct change to that of our public branch libraries. This along with the number of teams and staff members responsible for different areas of work, the volume of meetings and acronyms, meant it all felt a bit overwhelming.

However, we started to feel comfortable quickly by attending induction events, work shadowing, training and regular welcome meetings. This, along with our own drive and determination, helped us to quickly feel settled.

The role feels like a good fit for us now and, after just under two years in post, we are very happy with the work that we do. We both felt welcomed and part of the Library Collections team almost immediately which was positively reinforced by comments from colleagues both inside and outside of our team. We certainly both feel that the tentacles of the Metadata team spread a lot further across the University than we expected them to. There have been so many opportunities for us to get involved with different work in addition to the more traditional cataloguing which has been brilliant for our personal development. We don’t think this would have been the case had we not been as open to change and as enthusiastic as we have both tried to be.

We believe that our positivity, enthusiasm, initiative and drive were all contributory factors to our being offered the jobs. We have not undertaken formal education to embark on the roles of Metadata Assistants. This naturally meant that our arrival brought with it a different point of view due to our public library backgrounds. This has been a significant factor in this role as, due to the high number of adjustments we experienced working for public libraries, we are not afraid to ask questions and change processes when needed. This
sometimes sparks debate but often leads to positive results.

We endeavour to consider how we CAN do things, rather than presuming we can’t because we haven’t done it before. Subsequently, we have made a number of positive impacts during our short time working with metadata. Our shared enthusiasm has helped towards building an effective team. Our jobs could have been a lot more insular if we had allowed it, but as we were both completely new to this type of role we were able to mould them into what we felt they should be.

We have made some steps towards changing the face of the Metadata team and have diluted the stereotype often associated with the role. We have achieved this by completing a number of successful projects which have frequently led to our liaising with other teams both inside and out of our department. Good working relationships have developed with other teams, particularly within our specialisms. Clair works with Academic Liaison Services on the Kent Academic Repository (KAR), Article Processing Charges (APCs) and providing research support whilst Josie works with the special collections, which can include additional liaison with Canterbury Cathedral Library & Archives.

Cross-team working also enables us to work beyond our specialisms. This has been a substantial influence on the improved perception of the Metadata team and is a key reason why barriers that made the team insular have now been eradicated. Consequently, we have similarly demonstrated that we are capable of more than cataloguing and have improved the approachability of the team as a whole. This is evident through the increased number of requests we have received for work shadowing and regular meetings now held with other teams that are today part of everyday Metadata life. We have been approached to work on various cross-team projects, some of which are ongoing and these enable us to continually engage with other teams. These have included whole library reclassification, relocation of British Government publications, managing e-theses, implementation of new systems and ongoing stock moves. Projects are where we can continue to make a vital and positive impact and, along with improved communication and working practices, prevents us from working in a bubble.

However, this hasn’t been without its challenges. Due to the different areas of responsibility we have to ensure we manage our time effectively. We find that using our online calendars really helps with this so we can block out patches of time for each task. We also have a team rota for our shared responsibilities to ensure that one member of the team doesn’t get given the lion’s share of the day-to-day work, such as general cataloguing queries. We also make sure our duties are covered when we go on leave or are out of the office, and we have a produced an online wiki where procedures for our main duties are recorded, so that work can be picked up if needs be when we are away.

The role has been a bit of a ‘baptism by fire’ from the beginning, with us having to learn a lot rather quickly.
This is something we have both taken in our stride and have enjoyed the challenge of, but we have to be careful that we take the time to understand what we are doing properly so as not to cause errors to arise down the line; accuracy is an integral part of our jobs!

As previously mentioned, Metadata has been a fast growing team at the University, growing from 1 to 5 members in a matter of 18 months. All four new members of the team did not come from metadata backgrounds and therefore required quite a bit of training along the way. This has obviously meant we have had to devote some of our time to supporting new members of the team, but this has also given us the opportunity to develop practices together and to give our procedures a fresh look to ensure they are still effective and efficient. We have also had some changes in direct line management along the way, which whilst not being detrimental to our work, has meant we have needed to use our initiative and be self-sufficient at times. Throughout all these changes though it’s been clear that we are very supported by our managers and that they have confidence in our decisions and practices.

Standards and policies are always changing and we try to make sure we keep up with this. Josie organised some RDA training for us in April this year, delivered by Bibliographic Data Services, and we keep our eye out for relevant training and conferences we can attend to support our development.

Communication can be an issue in such a large organisation. We strive to mitigate this by being transparent in our working practices and by involving other teams in any work we do which will affect them, as this can help guide us on how to move forward. However, we know we could improve communication within our own team by sharing information relating to our own work more frequently.

“\textit{When the winds of change blow, some people build walls and others build windmills}”

\textit{Chinese proverb}
So what’s next for us as Metadata Assistants? Well, we are both interested in undertaking some further education, whether that is a Higher Education course, such as a Master’s, or small chunks of learning, such as MOOCs. We are certainly keen to continue to develop our knowledge and enhance our skills, and our experience so far is that this is very supported at the University.

We feel that as a Metadata team, we should learn to SHOUT LOUDER about our achievements and the job which we do. We shouldn’t work in a bubble. Our experience of cross-team working has been a really positive one and has certainly helped raise the profile of the Metadata team. And we shouldn’t have to be a stereotype; metadata is an integral part of many aspects of Information Services and our roles should reflect that. We should celebrate the work we do.
Catalogue & Index is electronically published by the Cataloguing and Indexing Group of the Chartered Institute of Library and Information Professionals (CILIP) (Charity No. 313014)

Subscription rates: free to members of CIG; GBP 15.00 for non-members & institutions

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: Helen Garner, (h.j.garner@shu.ac.uk) and Karen Pierce (PierceKF@Cardiff.ac.uk).

For book reviews, please contact the Book Reviews Editor: Neil T. Nicholson, Cataloguing & Metadata Services Team Leader, National Library of Scotland, e: n.nicholson@nls.uk

ISSN 0008-7629

CIG website: http://www.cilip.org.uk/cataloguing-and-indexing-group/catalogue-index

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