Minutes of the Cataloguing and Authorities User Group Meeting  
Monday, 15 September 2014  
Alice Moulton Room, Gerstein Library  
10:00 – 11:30 am

Present:  
James Mason (Music), Elisa Sze (Inforum), P. J. Carefoote (Fisher), Anne McGillivray (Knox), Mary Reynolds (Regis), Carmen Socknat (Vic), Mary Jaques (St Augustine’s), Jeff Newman (New), Ravid David (Scholars Portal Robarts), Marlene van Ballegooie (Robarts), Jordan Hale (Map & Data), Irene Wu (Rom), H. Rashid (Law), Lisa Andrews-Attwater (Rotman), Susan Williams (UTM), Joan Links (Media Commons), P.J. MacDougall (Massey), Juliya Borie (Robarts), Lana Soglasnova (Robarts), Kate MacDonald, Chair, (Trinity), Bilal Khalid (ITS), Alastair Boyd (Robarts).

Regrets: Michael Bramah (SMC)

1. Agenda approved.


3. ‘Linked Data’ summit summary (Alastair Boyd):

In conjunction with this year’s Access conference held in Toronto, representatives from the 5 largest Canadian academic libraries (U of T, McGill, UdeM, University of Alberta, UBC) met to discuss and learn about linked data projects, and how we might develop a Canadian linked data initiative. (Funding for such collaboration remains an issue.)

- **Report from Stanford University (Phil Schreur):**
  - Stanford is forging ahead with the move to Linked Data, as a parallel process to MARC format cataloguing in Sirsi, in partnership with Cornell and Harvard. The idea is more than a proof of concept: it is to create a pool of “native” library linked data, not just conversions from MARC. For some formats, e.g. performed music, conventional MARC cataloguing leads to imprecise retrievals, because subject or form terminology in a MARC record applies to the record as a whole & cannot be directly linked with individual works contained in a sound recording.
  - Each partner is concentrating on its own particular materials (e.g. Stanford is concentrating on music recordings to start with).

- **Report from Library of Congress (Beacher Wiggins and Sally McCallum):**
  - LC has just finished training for a pilot project to do cataloguing with Bibframe. Will involve 40 cataloguers, and last for 2-3 months. It too is a parallel process, because of LC’s obligation to distribute cataloguing in MARC 21 format.
    - Bibframe is a repackaging of data but “the data is still the data” and rules will still be needed for linking of data to happen.
    - Pilot is for creating native Bibframe data using the B-frame editor; a way of submitting the editor and the cataloguing process to iterative development
    - Involves all kinds of material (print, AV, music, recordings, maps)
  - So far only [!] 13 million out of 19 million legacy records have been converted.
The pilot will reveal how cataloguers respond to Bibframe’s description levels (Work, Instance, Item). They have customized the Bibframe editor extensively; has “type-ahead” and drop-down list features. The pilot will test how well those work. It uses RDA terminology, and has direct links to rules. LC will make this available to other libraries.

The hope is that the pilot will answer various questions about production level cataloguing with Bibframe: how does searching for “records” work? It is a new model for using data and inputting data.

The pilot is NOT doing the following:
- No development of end-user interfaces and access to bibframe data (i.e. not building a new ILS)
- No acquisitions information
- No working with Holdings information in the Bibframe model
- Distribution is still through MARC for many years to come
- LC will make training material available as well as their version of the editor.

**Report from OCLC (Roy Tennant and John Chapman):**
OCLC has already created “work IDs” (197 million) — entities with URIs that aggregate links connected with Works, including subjects, descriptions, etc. An example is at [http://worldcat.org/entity/work/id/463752](http://worldcat.org/entity/work/id/463752). They are just about to do a similar project for people; “Persons” entities (18 million in progress). These too will aggregate things like ISNIs, ORCID identifiers, links to VIAF, etc., to provide a linkable entity through which information by and about people can be linked to.

OCLC has chosen schema.org structures and labels with “bibliographic extensions” (see [bib.schema.org](http://bib.schema.org)). These Work and Person entities are a pilot for new services available to OCLC members.

It is hoped that preliminary training for cataloguers here at UofT will occur at some point during the current academic year.

**4. Endeca Update (Bilal Khalid, ITS)**
- A new format designation: Added to catalogue in response to a request from OISE. The term ‘object’ will now represent realia, kits, educational toys, etc.

This will not require new coding on the part of cataloguers as the information is currently taken from the GMD. The intent is to improve the public access display of these types of records based on the information that is already inputted by cataloguers.

Example search: [https://onesearch.library.utoronto.ca/onesearch/dominoes%20kit/](https://onesearch.library.utoronto.ca/onesearch/dominoes%20kit/)

- License information: working to include information about licensing restrictions within catalogue records for online resources to show the number of concurrent users and how many licenses are available.
- Continuing to work on improved display of MARC records, especially in Internet Explorer.
- UTF8 Characters: should now display correctly in the MARC section of the records in the OPAC.
• ITS has an ongoing maintenance schedule for fixing problems; the introduction of new features are scheduled separately.

• Currently working on the re-introduction of the ‘shelf browse’ feature. It is hoped that this will be finished and live by Friday, 18 September.

• Working to improve the known item search capability in order to assist with the discovery of known items, improve the relevance ranking of items found, and improve their display.

• Working on search autocomplete functionality for known items.

• At present, if a record has more than 100 items attached, anything above 100 will not display in Endeca, giving the mistaken impression that there are only 100 attached items when there could well be more. Fixing this problem requires interaction with Sirsi, but once this is resolved the full record should display fairly soon.

Example record:

Number of libraries with holdings visible in the online catalogue: 3
Actual number of libraries with holdings when looking at same record in Sirsi: 6

• The same issue is being dealt with in terms of displaying items on order.
• With regards to RDA, work is being done to begin displaying some RDA fields.
• The current time lag from creation of record to its appearance in Endeca is about 15 minutes.
• Kate MacDonald asked about the refine option for catalogue search results and its display on mobile devices, particularly iPhones since ‘the more libraries’ function can be hidden and it can be difficult to see the holdings for a specific library. Bilal said he would check into this [update: Kate has submitted screen shots to Bilal].

5. RDA versus AACR2 versus ‘hybrid’ records (Alastair Boyd)

i. There has been some further debate about GMDs, and the presence or absence of the 245 subfield |h, depending on whether records were constructed according to AACR2 or RDA; confusing because all OCLC WorldCat legacy records have been converted into “hybrid” records. But even these hybrid records contain the RDA-style MARC fields for content, medium and carrier —

- **336 - Content Type** e.g. performed music$2rdaccontent
- **337 - Media Type** e.g. audio$2rdamedia
- **338 - Carrier Type** e.g. audio disc$2rdacarrier

— rather than GMDs. These fields between them convey the same information as the GMD used to, in a more elaborate but consistent way. (GMDs were a mixture of content and medium: e.g. *game* versus *microform*). Therefore, RDA and hybrid records with these fields should not have a GMD inserted into them.
Endeca wish list: instead of a text string ([sound recording], [electronic resource]): best way to make print versus non-print items visually distinct in Endeca hitlists would be some kind of icon, such as you find in WorldCat.

These icons+labels are determined by record coding, just like the Endeca format facets. If/when this enhanced display is done, we could switch off the display of the 245 subfield |h which is no longer a reliable flag for non-print resources.

ii. RDA field for publishing information: Alastair provided a follow-up on this topic that had come up at a previous meeting when he thought that Endeca now recognized the RDA field for publishing information, the repeatable 264 field and cataloguers could stop editing 264s back into 260s. However, a couple of people pointed out over the summer that imprint information was missing from online catalogue and upon examination it was for RDA records with 264 fields. Therefore, in the short-term, Marlene and Maks will resume the batch conversions of 264 fields back into 260s. We hope that it won’t be too long before Endeca recognizes 264 fields again, and displays their contents.

iii. Also on the Endeca wish-list: to suppress the display, as hyperlinked search terms, of 65X headings that are not LCSH or MeSH (such as FAST headings, or French and German terms). It clutters up the display and produces very misleading results if users try follow-on searches using these links.

iv. Turning from Endeca to Sirsi: Wenran has installed a Sirsi patch that has fixed the Workflows browse bug that produced such weird browse lists (e.g. “eglise l’oeccumenisme” that showed you titles starting with “Egy”). The searches now appear to be working properly.
6. RDA Authorities Guide (Juliya Borie)

The tutorial that Juliya created to accompany training sessions on helping cataloguers form authority records according to RDA can now be found on the Cataloguing Resource Page → Policies & Guidelines → Authority. The tutorial may be found both in video and printed formats, and more information is available on the FAQ sheet as well.

http://guides.library.utoronto.ca/c.php?g=251371&p=1674010

7. Questions

Lana Soglasnova asked Bilal if it might be possible to specify language of search in Endeca to improve accurate hits. The programme defaults to English and returns hits that don’t reflect the original search in another language. Bilal answered that they will do their best to ensure the full slate of holdings are returned, especially by doing proper searches of the 8xx fields, for example.

Alastair asked about the pool of pre-catalogued materials we have whose home location displays as ‘In process – may be requested’. The problem is there appears to be no mechanism to actually request them. Is this an interface or programming problem that can be fixed? Bilal confirmed that he would look into it, but that Marc Lalonde was the main contact for that project.