

## Minutes of the Cataloguing and Authorities Users' Group Meeting September 21, 2009, Robarts Library Room 4049

*Present:* Lisa Andrews-Attwater (*Rotman*), Alastair Boyd (*Robarts*), Michael Bramah (*SMC*), Mary Canning (*UTM*), Bernie Disonglo (*Robarts*), Anne Dondertman (*Fisher*), Heng Ge (*CHKL*), Anna Golodnitsky (*Knox*), Monica Hypher (*CIRHR*), Blair Kuntz (*Robarts*), Diana Liang (*UTSC*), Joan Links (*Media Commons*), Kate MacDonald (*Trinity*), P.J. MacDougall (*Massey*), James Mason (*Music*), Mary McDiarmid (*Baycrest/HSICT*), Stephen Qiao (*East Asian*), H. Rashid (*Law*), Mary Reynolds (*Regis*), Mary Ruscillo (*Robarts*), Anna Slawek (*Robarts*), Sherry Smugler (*Gov Pubs*), Carmen Socknat (*Victoria*), Eva Spevak (*Robarts*), Elisa Sze (*Inforum*), Marlene van Ballegooie (*Robarts*), Irene Wu (*ROM*)

### 1. Minutes and Business arising from March 9 Meeting

Minutes approved as written. No business arising.

### 2. Report from the Metadata Librarian

Marlene van Ballegooie reported on the following projects:

- *OCLC WorldCat error report and record fixes*  
Of the initial UTL records extracted and submitted to OCLC there were 92,969 records which failed to meet the minimal requirements to be entered into the WorldCat database. Within this record set, a total of 170,814 errors were reported. These ranged from invalid subfield codes in the **245** and **300** fields, to invalid **008** fixed field codes (e.g. country and language codes), to invalid **007** control fields. By parsing the OCLC error reports and converting to a database, Marlene has been able to identify and fix a large number of errors programmatically.
- *Form Subdivisions project*  
Following up the discussion from the last CAUG meeting, Marlene has identified numerous form subdivisions incorrectly coded as topical subdivisions in our catalogue (i.e. subfield **|x** instead of **|v**). Such inconsistencies in coding have become highly visible in the new Endeca catalogue, where topics and forms/genres are separate facets. So this retrospective conversion project used scripts to flip the incorrect codes from **|x** to **|v**. One complication: many form subdivisions can also be topical subdivisions. For example, Librettos can be coded as **|x** for an item *about* librettos and it can be coded as **|v** for items that *are* librettos. So, first we will concentrate on those subdivisions that can only be forms, and those for which coding elsewhere in the record can be used to verify that the subdivision should be a form and not a subject. Work with the form subdivisions will continue as time permits.
- *MARC 041 Language Codes*  
Endeca looks for language codes both in the **008** fixed fields and in the **041** field. When searching the **041**, Endeca is looking for a 3-character language code. If there are more than 3 characters in a string, Endeca disregards the data because it can't make a match with the MARC code list for languages. The current rule for encoding language information for items with multiple languages is that each language of the text is coded in a separate subfield **|a**, but before 2001, it was the practice to string multiple language codes together in one subfield. For example, a string previously coded as **|aengfreger** is now coded instead as **|aeng|afr|a|ger** and it is this later form that Endeca expects to find. Because so many of our records were created before 2001 and contain single **|a** fields with multiple language codes, Endeca has been disregarding a lot of language information.

Marlene has dumped out and recoded around 442,000 **041** fields, which will be written back into Sirsi when some ITS programming time is available.

- *Update on Genre/Form authority records in Sirsi*  
Library of Congress Genre/Form authority records are now appearing in our authority file as we load new batches of subject authority files. However, Sirsi is not currently set up to validate against these authorities or even let us search for them except by LCCN. Marlene is working on the necessary reconfigurations in Workflows, such as creating a new “Genre” index, to allow us to make proper use of these new authorities.
- *Metadata conversions and modifications*  
Marlene’s other ongoing projects include: vetting MARC records from e-book vendors; working with T-Space, generating MARC records for theses, and also generating Dublin Core records for U of T’s exam collection which will also be housed in T-Space this fall; and working on generating MARC records for items held in locally-created digital collections.

### **3. Updates**

#### *Thesis cataloguing*

Starting with this Fall Convocation, all theses will be submitted to SGS in electronic format only. Since print copies will no longer arrive in any cataloguing department, a new procedure is being worked out. Marlene has a script (as mentioned in her report) that will generate MARC records from the T-Space Dublin Core metadata and load them into Sirsi. These will be equivalent to the records we have been getting from ProQuest. Libraries that currently provide fuller cataloguing for theses no longer need to create their own record first; they can upgrade these script-generated records with subject headings and other controlled access points. Because this is all so new, there are still some uncertainties about the timing of the various steps in the process. In theory, T-Space thesis records are made public two weeks after convocation, but in practice it seems to be taking much longer: two months has been suggested as a more realistic estimate.

#### *Endeca*

There remains quite a long “To Do” list for fixes and upgrades to the new catalogue interface. Unfortunately there has been an interruption in Endeca development, because ITS has been working on setting up a Media Server for streaming or downloading lectures. This is part of the University’s pandemic preparations and so has been the top priority. ITS was very close to finishing the new Endeca browse indexes when work was suspended, so this feature should be ready some time next month.

#### *RDA planning*

Like many other libraries, the U of T central library is waiting for the outcome of the current RDA testing being carried out by LC, NLM, Library and Archives Canada, and a few other large libraries. There will be a six month pilot project — three months of RDA training, three months of creating catalogue records using RDA — after which LC and the other US national libraries will spend a further three months analyzing the results and listening to feedback from other libraries. They will then decide on the operational, technical, and economic feasibility of implementing RDA. (See <http://www.loc.gov/bibliographic-future/rda/> for details). This means that a decision from LC is unlikely to be announced before mid-summer 2010. And even if the decision is in favour of RDA, it is likely that implementation will be phased in gradually. Thus there will be sufficient time to set up training, make modifications to Sirsi, etc., even if we wait until next summer’s decision.

#### **4. Other business**

- Alastair asked how many libraries were sending back issues of periodicals to Downsview, while keeping current issues in the stacks or in a periodicals reading room. There is a problem in the catalogue where the Downsview holdings are displaying first in hitlists, which suggests to catalogue users that all volumes are offsite. This happens because each Downsview volume has a call number and item record, while current issues generally have only MARC holdings. In addition, the MARC holdings are attached to the owning library, so the catalogue will say “Library Has” even when partial holdings are at Downsview. This is another consequence of setting up Downsview as a library rather than a location. Catalogue users who view a single serial record in the catalogue will see conflicting information: the MARC holdings describes the volumes as being at the owning library in the normal home location, while immediately following is a list (often very long) of volumes at Downsview. Once more libraries start sending back issues to offsite storage it will become important to decide on a consistent strategy for overcoming these discrepancies.
- Alastair also announced that the next CAUG meeting will be his last as Chair. He has held the position since December 2002. Nominations are open for a successor to be elected at the next meeting, who will take over in 2010.

The next CAUG meeting will be in November. Precise date to be announced.

*The meeting adjourned at 11:30.*