**Authority control and the digital repository: what happens to controlled vocabulary once it's outside the ILS?**

Karen D. Miller

Monographic/Digital Projects Cataloger

Bibliographic Services Dept., Northwestern University Library, Evanston, IL

k-miller3@northwestern.edu

* Catalogers are accustomed to using controlled vocabularies with MARC catalog records. The ILS usually provides some form of authority control, generally providing validation that a heading is in the authority file and then updating that heading in the authority file and any bibliographic records that contain it if it changes.
* What happens when controlled vocabulary ventures outside the ILS?
* This happens in at least two ways:
	+ Repurposing of MARC data – MARC data crosswalked to other metadata schema for use on the web or in a digital repository.
	+ Creation of metadata outside the ILS – non-MARC data that takes advantage of controlled vocabulary. This might be created in Archon, Archivists Toolkit, ContentDM, or in an XML editor, "by hand".

*Lifecycle of a bibliographic record in an ILS*

* Catalogers create bibliographic records using descriptive fields and controlled vocabulary terms. The ILS provides a means of verifying that the controlled terms used are in the authority file and that they're coded correctly.
* Eventually, changes will be made to these records. Catalogers fix typos, add notes, etc. Authority records change when dates of death are added or preferred subject terms change. The ILS usually provides a means of updating associated bibliographic records when changes to the authority file are made.
* In the ILS, changes to bibliographic records are reflected immediately in the OPAC, so patrons get the benefit of updated authority records.
* The lifecycle of a bibliographic record in the ILS is self contained, assuming the ILS provides some means of updating authority records.
* Many libraries catalog digital resources in their ILS and store the digital objects themselves in their repository. In order to preserve descriptive metadata along with the digital resource, they simply crosswalk their catalog record into an XML format and store it in the digital repository.
* The digital repository is a *repository* by definition – it isn't a cataloging tool and it doesn't provide authority control. So what happens when a change is made to the record in the ILS or an authorized heading is changed? The record in the ILS is updated, but the record in the digital repository is not.
* In the past 12 months, less than half of a percent of authorized headings in the LCSH were changed. This is a small percentage, but the numbers are large and, particularly when it comes to adding dates of death, these are changes we want to capture.
* Changes to authorized headings (1xx fields) in LCSH and LC/NACO between March 2009 and February 2010 (12 months):
	+ LCSH: 384,393 headings, 1,663 changed (0.43%)
	+ LC/NACO: 7,715,612 headings, 32,033 changed (0.42%)

*Outside the ILS:*

* There are other means of creating metadata in XML – Archivist's Toolkit, which creates EAD; VCat, which creates VRA Core, and ContentDM. In addition, there is a growing number of locally developed editors such as Duke's Trident, Indiana's PhotoCat, Brown's XForm MODS editor, and other editors developed at the National Science Digital Library (NSDL), University of Virginia, University of Alberta, and others. Most of these editors provide some form of controlled vocabularies and authority control.
* In most cases these editors are like the ILS in that they are used to create metadata records that are then exported and stored in the digital repository.
* We still can expect the controlled vocabularies to change after the records have been stored in the digital repository and we can expect descriptive aspects of the records to change as well.
* These editors might be in a better position to keep in touch with their metadata records once they've left the nest. Most digital repositories will have an automated means of sending a finished record from the editor to the repository. This could be adjusted to send any changed record from the editor to the repository, as long as an update to the authority file registers as a change.
* If you're creating records by hand or with an editor that has no authority control, then you're back to square one – no automated way to re-generate the records when they've changed.

*What you can do:*

* Identify a database of record. Records should be edited in only one database, say the ILS or the metadata editor, and then exported to the digital repository. Do not make changes in both places – this is a waste of time! Consider exported records in the digital repository to be "read only" – don't edit them after export or you'll lose the changes when you need to re-export.
* Automate the transfer of records from the database of record into the digital repository. Build in a means of recording changes to records and trigger the automated transfer when a change is made.
* If you're encoding by hand or using an editor without authority control try to capture a control number in the XML so that you can identify changed controlled vocabulary later. We haven't solved this one yet!

*Other solutions:*

* Handles : use a handles to avoid changing URLs
* OAI-PMH : OAI harvesting allows for the identification of modified, deleted, or new records, so a harvester can delete or replace records already harvested.
* Linked data?