

A Cataloging Research Blitz !!!

*Celebrating 2010:
The Year of Cataloging Research*
APRIL 19, 2010

COORDINATED BY:

Jane Greenberg, Professor and Director of the
SILS Metadata Research Center

Margaretta Yarborough, Head, Resource
Description & Management Dept., UNC Libraries

Renee McBride, Head, Special Formats &
Metadata Section, Resource Description &
Management Dept., UNC Libraries



UNC
SCHOOL OF INFORMATION
AND LIBRARY SCIENCE

Metadata Research Center <MRC>



UNC
UNIVERSITY LIBRARIES

It hurts...



- “Today, if you have an Internet connection, you have at your fingertips an amount of information previously available only to those with access to the world’s greatest libraries – indeed, in most respects what is available through the Internet dwarfs those libraries, and it is incomparably easier to find what you need.”
- Peter Singer, Professor, Princeton University
- <http://www.project-syndicate.org/commentary/singer61/English>
- NGC4LIB list, Sat. April 17, 2010

ALL RIGHTS RESERVED
<http://www.cartoonbank.com>



"Never, ever, think outside the box."



Bringing Authority Back to a Faceted Browse Catalog

1. UNC Libraries are trying to put authority data back in the path of researchers using the Endeca catalog.
2. Real-time auto-suggest for authors, titles, and subject headings as the user types; Worldcat identites NameFinder service (Search TRLN)
3. Number of subject searches increased nearly 200%; Author searches increased by nearly 70%.

Beyond Ranked Lists: Users and Faceted Library Catalog Searching



1. How people are using facets offered to them by the library catalog?
2. Transaction log analysis/ User studies
 - For UNC library catalog, 6% total operations are facet operations
 - Users cluster into certain categories of searching behavior
 - Transition matrix shows the pattern of searching behavior



A Subtle Censorship: A Study of the Shelving Practices of Young Adult Literature in North Carolina

1. The study primarily focuses on purposeful or inadvertent “censorship by categorization”
2. Examination of library OPAC, searching for placement of 10 controversial young adult titles
3. Study found a lack of consistency in placement worthy of further evaluation and/or a change in the way materials are placed

Into the Future: FRBR, the Semantic Web, and Possible Applications for Music Cataloging



1. Examines the applications of FRBR and Semantic Web technology on traditional music cataloging.
2. Uses a detailed literature review and an exploratory study of practical models.
3. FRBR's ability to collocate items makes it an attractive option for music-oriented displays.
4. Migrating to XML-based authority and bibliographic records, applying controlled-vocabulary tools like SKOS to LCSHs and LCCNs, and merging FRBRized records with Semantic Web technology might allow music catalogers to begin harnessing the Semantic Web's many benefits.

WHAT? CONTENTdm Authority Control Project (authority control of names and subjects)

WHEN? Began Fall 2009

WHO?

- Lynn Holdzkom (Head, Technical Services, Special Collections)
- Renée McBride** (Head, Special Formats & Metadata Section, Resource Description & Management)
- Tim Shearer (Head, Applications Development Section, Library Systems)
- Jake Vaccaro (CALA)
- CONTENTdm collection managers and catalogers

WHY?

- Ensure consistency across CONTENTdm digital collections
- Offer end users more consistent results from browse lists
- Improve ability to manage controlled vocabularies
- Be prepared to add CONTENTdm collections to Endeca index

HOW?

- Employ shared controlled vocabularies (CV)
- Compare contents of CV with library catalog & LCNAF to determine correct forms of names
- Update CV with corrected forms of names
- Correct metadata to match corrected CV
- Meetings, workshops & regular communication with CONTENTdm collection managers
- Active maintenance & use of CONTENTdm wiki

WHERE?

Names: As of April 2010 Jake is on letter P; Renée is on G

Subjects: Hope to get started by Fall 2010



HIVE v. BioPortal: a Comparison of Two Annotators

- This research focused on automatic metadata generation tools, specifically automatic annotators.
- An experiment was conducted in which a document set was tested on both annotators to determine:
 - ▲ Precision
 - ▲ Specificity

	Precision	Specificity	Exhaustivity
BioPortal	0.012	2	1.7
HIVE	0.209	2.3	2.2



SPIRO-V: The Future of Multidisciplinary Control Vocabularies and Ontologies

1. Controlled Vocabularies for various domains – Need of the Hour
2. The SPIRO-V Model
 - Vocabulary Generation Techniques Used
 1. Manual
 2. Automatic
3. SPIRO-V Implementation so far
 1. Relational DB
 2. Web User Interface
 3. Expert User
1. Future implementation and the vision
 - Ontology – A way to access and visualize data



Cataloging Data: Pitfalls for GIS and Numeric Data Sets

- Problem: How to make GIS & numeric data resources more discoverable
- Method – Exploring:
 - Special metadata standards (FGDC & DDI)
 - Cooperative cataloging with Odum Inst./Duke
 - Integrating Acquisitions module with ILS
- Results so far:
 - Cataloging with special standards not practical
 - Control issues in shared cataloging
- Future Research:
 - Scalability of cataloging (faculty solution; data set to layer level cataloging or merely Dublin Core)
 - Copy cataloging from other standards (MERMAid)
 - Search interface suited to GIS/numeric data

GIS
|

#

Numeric data set:

a mass of raw numbers collected by a researcher;

may be re-used by another researcher for a new purpose



Extracting & Ingesting Data & Metadata, DL → iRODS

1. Whether and how you can move data/metadata from a publishing platform (e.g., data archive & digital library) to preservation data grid using existing standards, such as the OAI-PMH
2. Used XSLT transformation to move the DDI metadata to Attribute-Value-Unit triplets; used Jargon (an iRODS API) to transfer the data objects
3. Surprise: move over the metadata file as an object, too; hoping to make it a production system