

Minutes
Catalogers Group
July 24, 2003
2:30-3:30
West Electronic (Media) Classroom

Present: Valerie Bross, Beth Feinberg, Joan LoPear, Renee McBride, Luiz Mendes, Caroline Miller, Jeff Morehead, Nancy Norris, Hao Phan, Louise Ratliff, John Riemer, Angela Riggio, Rita Stumps (recorder)

I. **Report on ALA/CLA Annual Conference, June 19-25, 2003**
(Louise Ratliff)

ALCTS CCS CC:DA program: "Don't be dysfunctional : how to put the Functional Requirements for Bibliographic Records (FRBR) in your future."

Matthew Beacom for Barbara Tillett: Overview of FRBR

For the Power Point presentation and other CC:DA documents, see:

<http://www.ala.org/alcts/organization/ccs/ccda/ccda.html>

Another version located at

http://www.ala.org/Content/NavigationMenu/ALCTS/Continuing_Education2/Presentations/Tillett.ppt

Vinod Chachra, VTLS: Virtua and implementation of FRBR

See: <http://www.vtls.com/documents/FRBR.PPT>

[For convenience, files have been copied to the I: drive. Launch I:\VTLS Demo June 2003\FRBR.PPT]

Sally McCallum for Tom Delsey: "Mapping MARC to FRBR"

See: <http://www.loc.gov/marc/marc-functional-analysis/functional-analysis.html>

At the bottom of this page there is a link to the FRBR Display Tool. "The **FRBR Display Tool** takes the work done in "Displays for Multiple Versions from MARC 21 and FRBR" one step further. It transforms the bibliographic data found in MARC record files into meaningful displays by grouping the bibliographic data into the "Work," "Expression" and "Manifestation" FRBR concepts. The below matching and sorting specifications outline the processes used to generate the FRBR displays." A Web interface to this tool is "coming soon."

Glenn Patton, OCLC: "What can FRBR do for you?" See:

http://www.ala.org/Content/NavigationMenu/ALCTS/Continuing_Education2/Presentations/Patton-FRBR.ppt

Main point, "It's all about relationships," for example, earlier/later forms of names; performers/works performed.

Further reading: Denton, William, "FRBR and fundamental cataloging rules," May 2003.

Available at <http://www.miskatonic.org/library/frbr.html> (Note: this is a helpful discussion of FRBR, set in the context of the history and principles of cataloging – 13 p.)

Key points: FRBR is a conceptual model, not a data model. It facilitates finding and collocation, and creates clearer, more useful relationships.

Training for Effective Subject Cataloging: A Preview of the ALCTS-Program for Cooperative Cataloging (PCC) Initiative (see printed handouts)

Ana Cristán introduced the program, giving Carol Hixson's remarks <http://darkwing.uoregon.edu/~chixson/subj/cghRemarks.pdf> .

Bill Garrison presented one module, "Application of Subdivisions," from the new continuing-education course on subject cataloging. This course, based on the highly successful SCCTP model, will be presented as a pre-conference to the 2004 ALA Annual Conference in Orlando. It is designed to teach basic subject cataloging using LCSH, and will be a 2-day workshop. The major topics to be covered each day are:

Day 1

- Subject analysis principles
- Principles and structure of LCSH: Overview
- Tools for subject analysis: how to use them and their availability
- Application of LCSH: assigning and constructing subject headings (pt. 1)

Day 2

- Application of LCSH: assigning and constructing subject headings (pt. 2)
- MARC coding
- Evaluating existing headings and Checklist of steps
- Specialized subject areas and formats
- Introduction to SACO

Audience members were very enthusiastic about this course and gave Bill a hearty round of applause after his presentation of the segment on "Application of Subdivisions." For more information, see "Basic Subject Cataloging Using Library of Congress Subject Headings (an ALCTS/SAC-PCC/SCT Joint Initiative)" at <http://darkwing.uoregon.edu/~chixson/subj/publicsubj.html>

PCC Participants Meeting

Bob Wolven, Chair, convened the meeting and reported on the state of the PCC program and gave highlights of activities during the past 6 months. Full minutes are available at: <http://www.loc.gov/catdir/pcc/pccpart03a.html>

Highlights of note:

Wolven spoke about training initiatives and announced upcoming training sessions. He also noted international outreach efforts.

The PCC Participants' Meeting focused its attention on CONSER on the occasion of its 30th anniversary celebration, the international flavor of the program, and on Jean Hirons, CONSER Coordinator, on the occasion of her retirement. PCC participants were treated to exposés about the CONSER program and its training model (SCCTP) by Canadian colleagues.

Liz McKeen, Director of Bibliographic Access, Library and Archives of Canada (LAC), offered a historical perspective on LAC's involvement with the CONSER program.

Jean Hirons, CONSER Coordinator provided her perspective for future directions for the CONSER program and the challenges to be faced by her successor.

LITA Top Technology Trends

Eight panelists presented remarks about issues of current interest in information technology: Karen Coyle (CDL), Joan Frye Williams, Clifford Lynch (CNI), Walt Crawford (RLG), Thomas

Dowling (OhioLink), Elizabeth Lawly (DITech, Rochester), Tom Wilson (U. of Md.), Marshall Breeding (Vanderbilt).

Selective notes:

Coyle briefly stated that for display of MARC records, listing search results in alphabetical order needs to be replaced by something more meaningful. Also, we need to “dis-integrate” our library systems so we don’t have to use only one vendor for all functionality.

Williams talked about the issue of accommodating user-supplied equipment (e.g., PDAs) in libraries. We need to show expertise about *hardware* so that we are seen as the source of *content*.

Lynch remarked that wireless technology is deploying faster than expected, and that people are “packing” (carrying) storage. We need to know that people are amassing sizeable libraries of “stuff.” For example, students download quantities of journal articles and take them home after graduation. Also, e-scholarship in terms of digital humanities is emerging. Descriptive metadata is coming into serious concern.

Crawford noted the demise of dedicated e-book readers.

Dowling spoke about an email spam filter used in Mozilla 1.3 and above that really works: you identify messages that are and are not spam, and you tell the filter to learn the difference. Could this be applied to searching in a database?

Lawly commented on weblogs and said they are having more impact on information gathering and publishing. Weblogs enable publishing on the web as well as user comments on what is published. They are a source of alternative information.

(Note: from www.blogger.com, the definition of a blog is: A blog is a web page made up of usually short, frequently updated posts that are arranged chronologically—like a what’s new page or a journal. The content and purposes of blogs varies greatly—from links and commentary about other web sites, to news about a company/person/idea, to diaries, photos, poetry, mini-essays, project updates, even fiction.

Wilson spoke about accessibility and the need to simplify interfaces (e.g., small display size for use with various devices). He also mentioned the need to be involved in the standards-making process, since the technologies we use are not specific to the library.

“Cliff’s Notes” – Clifford Lynch (selective notes)

Clifford Lynch, director of the Coalition for Networked Information, gave a talk which focused on the need to recognize how the way we are doing science is changing. Scientists are doing data mining and they need facilities for data management, data mining, and data curation. Disciplinary data centers are needed, and libraries need to be involved. Each university should provide for the dissemination and stewardship of works produced at that university, in the form of institutional repositories.

This need exists not only for science but also for the humanities and social sciences. With the “rebirth of the critical edition,” there is a substantial revolution in the practice of disciplines, but there is no funding agency comparable to the National Science Foundation.

There are big policy issues involved with what universities want to be in the digital world. For example, who owns course materials or student work?

He commented about the prospects for personalization and digital libraries; a personal digital library is formed as you download or look at materials. There are issues of privacy and permanence.

On the topic of computational linguistics and data mining, he said that you can derive descriptive metadata by having the machine analyze digital content, especially the contents of corpora of image material. This will be important. Images are becoming more accessible, but the linking text is too weak to provide good finding ability; the ability to bridge is not good.

Speaking about blogs (weblogs), he remarked that it is a social phenomenon that fits in someplace, but it is not clear where.

Regarding digital stewardship, we are becoming dependent upon digital resources, and sharing is important. Digital backup of cultural resources is a good idea.

LITA President's Program: Universal Access to All Human Knowledge
Brewster Kahle, Digital Librarian and Co-founder of the Internet Archive

People are learning based on what they see on the internet. Access to all human knowledge is within our grasp. Should we enable it? – Yes. Can we provide all published materials? May we? – There are restrictions. Will we? He cannot answer that question.

Texts: How many books are there? He estimates 100 million (with 26 million in the Library of Congress). How big is this? At about 40 megabytes per book, it equals about 40 terabytes! Kahle built a bookmobile where you can print on demand materials which are out of copyright, at a cost of \$2 per book for printing and binding. In India, they will digitize the books for free! In India there are 30 bookmobiles, 2 in Egypt and 2 in Uganda. There are none in the U.S.

Audio: He asked bands whether he could host digital recordings of their concerts, and many said yes! These are 1-gigabyte sized files of good quality. He is now getting other genres of music, and has offered unlimited storage for free.

Moving images: the television archive records international broadcasts 24 hours per day, including news coverage.

Software: There are about 50,000 titles, but it is illegal to copy software so there is no way to include it in the digital archive.

WWW: Web sites have been archived beginning in 1996, and the archive is massive—500 terabytes and growing.

What should be done about preservation? Hard drives crash, programmers make errors. We have the responsibility for archiving internet resources, but he's not sure that we will do it.

Extracts from <http://archive.org> Louise highly recommends taking a look.

The Internet Archive Wayback Machine contains over 100 terabytes of data and is currently growing at a rate of 12 terabytes per month. This eclipses the amount of text contained in the world's largest libraries, including the Library of Congress. If you tried to place the entire contents of the archive onto floppy disks (we don't recommend this!) and laid them end to end, it would stretch from New York, past Los Angeles, and halfway to Hawaii.

Most societies place importance on preserving artifacts of their culture and heritage. Without such artifacts, civilization has no memory and no mechanism to learn from its successes and failures. Our culture now produces more and more artifacts in digital form. The Archive's mission is to help preserve those artifacts and create an Internet library for researchers, historians, and scholars. The Archive collaborates with institutions including the Library of Congress and the Smithsonian.

The Internet Archive is collaborating with various groups to provide valuable audio archives, including the highest quality live concerts, independent news programs, and much more.

Also: Louise visited the Endeavor booth at ALA. Saw demo of Unicode (Chinese, Hebrew, Arabic, OPAC views). Not yet in Beta testing; by end of summer, hopefully. Unicode should be in next Endeavor release. Louise thought it looked good.

II. NASIG Portland reports (Angela & Valerie)
Angela's notes:

NASIG
Portland OR
June 26-29, 2003
Brief Notes
(Angela Riggio)

Plenary session. Leigh Watson Healy, Outsell, Inc.
Content Industry Outlook 2003: Asteroids That Are Changing the Information Landscape

Outsell is in the business of analyzing content provider companies, as well as the content buyer markets and the end users. Healy claims that content is "raining down" on users these days.

IC is lagging behind the IT community. Since it is a market-driven environment, the market is changing what vendors do. Although the industry is shrinking, companies like Google, in the content business, is growing, although roles within are changing.

End users in the business and education communities are finding what they want on their own (or they think they are). Users prefer independent searching, and prefer to print out what they find online.

The Asteroids

- ? Everything is integrating—there is content bursting out from inside and outside the firewall. Information specialists are expected to create and deliver content to users' desktops, according to specific user needs
- ? Companies (Kroval, Ovid, Hoover's, CBSMarketwatch, Google, Yahoo, IBM WebFountain, Microsoft) are both IT and CT companies. They have been born that way, so they don't face the same challenges that traditional information providers must face
- ? Technology—it's there, but vendors can't use it! Vendors seem to be providing content based on bad business models, not on user-based models. This must change.
- ? Implications—now, tech companies are getting into the content provider business—what it will all boil down to in the end is TRUST. Who will the user trust?

Recommendations for Libraries

- ? Know your users, what they want, how they want to get info., how they use that info.
- ? Develop partnerships with the IT community
- ? Learn to understand business models—libraries must market themselves!
- ? Libraries must self-assess

Recommendations for Vendors

- ? Know what is being used
- ? Move to a wider range of technical models; utilize XML and Web services
- ? Develop new business models
- ? Present information objects, not aggregations of content
- ? Forge new partnerships with information professionals

Serial Aggregations—Multiple Versions and the Virtual Union Catalog: The Combined Digital Library, SUNY and ExLibris Experiences

This presentation featured Michael Kaplan from ExLibris, who talked about the importance of keeping separate records for e-versions. This eases the updating process, and comes closer to

the FRBR model, etc. He demonstrated the 'perfect' user OPAC display, where records are not merged, but are viewed upon by demand—CDL!!!

Maggie Horn, Heather ? from SUNY consortium each gave their takes on the difficulties involved with the ExLibris migration. Timothy Gatti, head of cataloging at SUNY-Albany then described the way his department solved the problem created by enormous Web lists of e-journals by cataloging all of them, making them accessible instead via the library's OPAC (what a concept).

Case Studies in E-Serials Cataloging (Steve Shadle)

Shadle talked about the aggregator-neutral record during the first half of the workshop, and then moved to tricky examples of e-serials and IRs.

Cataloging Node, featuring Jean Hirons' GRBR presentation [a takeoff on FRBR]

Digital Preservation Conundrum (Abby Smith, CLIR; Eileen Fenton, JSTOR)

Smith described the paradoxes related to digital preservation:

1. Content—it resists fixation in the electronic world—quite different than paper. E-information is dynamic. How do we preserve content on Amazon.com, or GIS information, things that are generated on the fly? Who is responsible for fixing and maintaining (preserving) broken reference links contained within these objects?
2. Organizational—libraries license, they don't own. Copyright owners own, and most of them are not aware of preservation issues. Publishers, aggregators, etc. are also not in the preservation business, and there is little incentive for these parties to become interested in it.

Legally, libraries don't reserve the right to preserve. Libraries should partner with the owners of content. After materials have enjoyed their heyday, they should be turned over to organizations that would preserve them.

3. Users—we used to know who they were! We no longer know who they are, or if the materials we provide are adequate

Institutional responses: Repository efforts made by MIT, UC, Harvard, Yale, etc.); partnerships being formed by LC; 3rd party organizations like JSTOR.

Fenton mentioned JSTOR's plan to incorporate born-digital materials into their repository. Work is currently underway, but JSTOR must first examine the WHYS, the motivations, the monetary investments, etc. etc. etc. She continued with some basic definitions, and some basic assumptions about why preservation is important.

Hot Topics—Mergers/Consolidation/the-end-of-publishing-as-we-know-it (same as the session Valerie attended, but on a different day—seems like the discussion took a different course).

Each guest speaker received five minutes to pontificate:

Keith Courtney, Taylor & Francis: big biz only "buys what is for sale"—this isn't their fault!
Miriam Gilbert, Marcel Dekker: as a small family publisher, they get to know their authors and the content they are selling. They must learn from their customers, it's a "tough love" situation.

Michael Markwith, TDNet: Three questions can summarize this situation (questions that Markwith claims appear on his desktop every day~~s~~):

"Does size matter?" Yes, it provides resources to provide what is needed. "You need to consolidate now." Technology is driving the changes, it BREEDS consolidation. "Are you looking for a partner?" Look for ways to set up partnerships. Smaller publishers should look to relevant companies to meet changing needs—they need to keep their options open.

Kim Maxwell, MIT: at MIT, most subscription agents have changed over the last 5 years—what does this mean? Convergence is happening! This is a difficult process; there is great concern

about money, etc. There is a need for performance standards and communication among the vendor community.

The floor was then opened for discussion. Several comments from the floor were taken with some debate.

Valerie's notes:

NASIG 2003
June 26-29, 2003

NOTE: All handouts and reports for the 2003 sessions will be posted to the NASIG web site at: <http://www.nasig.org/> [but they are not on the site yet]

Sessions attended:

Friday

1. Keynote: Trends in the Information Content Industry (Leigh Watson Healy)
2. The Digital Preservation Conundrum (Eileen Fenton-JSTOR; Abby Smith-CLIR)
3. The Aggregator-Neutral Record & Case Studies in E-Serials Cataloging (Steve Shadle)
4. Cataloging Node (Jean Hirons, Christine Bellinger, Mavis Molto, Lai-Ying Hsiung)

Saturday

5. How are E-Journals Changing the Pattern of Use (Carol Tenopir, Donald W. King, Peter Boyce)
6. Usage Statistics: Taking E-Metrics to the Next Level (Oliver Pesch, EBSCO)
7. Starting with an Empty Map: Time/cost Assessment for Technical Services (Nancy Slight-Givney & Mary Greci)

Sunday

8. Hot Topics: Serials Consolidation (Keith Courtney-Taylor & Francis, Michael Markwith-TDNet, Kim Maxwell-MIT)
9. Keynote: Wrap-up Panel of Presenters (Jill Emery-U of Houston; Rick Anderson-Univ. of Nevada; Adam Chesler-Ingenta; Joan Conger-U of GA; Ted Fons-III)
10. Endeavor Voyager (facilitators Bob Persing-U of Penn; Maggie Rioux-Woods Hole)

Summaries:

1. Trends in the Information Content Industry (Leigh Watson Healy-Outsell, Inc)

Healy summarized research results completed by her company.

Market-side: Based on 100+ companies tracked since 1996-present, 50% of the growth is due to company acquisitions rather than new content.

Consumer-side: Polled 30,000 end users. The survey revealed that these users generally prefer self-serve than mediated service. Of those polled, 78% were pretty happy with what they found. 45% just use the open web rather than library services; but it is unclear whether they realized when they were using freely available v. library resources. While the Internet is preferred for searching, print is still preferred for reading; most users print out what they need.

On the horizon: Healy reported on the advent "Buzz-tracking" software. For a company, this software can track "all the buzz about you on the web" or "all the buzz about a specific

person/company you want to acquire." For example, some recording companies track the web for fan-groups, to help discover new artists.

In terms of content providers, Healy noted the following types of providers:

- o context-based delivery (for a specific subject niche)
- o repository-based products (based on XML)
- o products that permit increasingly granular uses of data
- o products from those new to content-provision: Microsoft, Adobe, Blackboard, SAP

2. The Digital Preservation Conundrum (Eileen Fenton-JSTOR; Abby Smith-CLIR)

Largely about the social context.

3. The Aggregator-Neutral Record & Case Studies in E-Serials Cataloging (Steve Shadle)

In the first segment, Shadle reviewed the fields for the Aggregator Neutral Record; the guidelines have been posted to the CONSER web site. Catalogers will need to be careful about cataloging from example, since most of the examples currently in OCLC will reflect OCLC merging strategies rather than application of the guidelines. The second segment consisted of discussion of a few case studies.

4. Cataloging Node (Jean Hirons, Christine Bellinger, Mavis Molto, Lai-Ying Hsiung)

Jean Hirons provided a perfectly hilarious parody of FRBR, called GRBR (pronounced gerber). This is a classic!

Bellinger, Molto, and Hsiung spoke about maintenance of records for e-journals. Bellinger & Molto use Serials Solutions. Lai-Ying discussed the use of Lexis-Nexis information, and the need for better techniques to pull together metadata in our local catalogs as is done in Melvyl--conserving the original metadata records, but indexing & presenting the unique information.

5. How are E-Journals Changing the Pattern of Use (Carol Tenopir, Donald W. King, Peter Boyce)

This group of people cooperated in a series of studies on e-journal use. They identified 3 stages of e-resource use:

- | | |
|------------|--|
| 1990-1995: | Early use: largely CD-ROM & print; a few e-journals |
| 2000- | Evolving: Mixed environment, with 40% of the journals online |
| 2001- | Advanced: Exemplified by the situation for astronomers (80% of the information is found on the web); the research environment is designed specifically for the way researchers work. Boyce emphasized the importance of developing a system for serving up information in ways that researchers need, rather than simply throwing e-resources up on the web. |

More information on this topic available at: <http://web.utk.edu/~tenopir/research/ejresearch.html>

6. Usage Statistics: Taking E-Metrics to the Next Level (Oliver Pesch, EBSCO)

Pesch discussed COUNTER, an emerging standard for counting statistics of e-resource use. The goal is to measure usage in a way that will serve various communities: publishers (to track usage royalties), libraries (to track whether a resource is used). Pesch described the challenges of defining things like: "What needs to be counted?" and "What does the tally really mean?"

7. Starting with an Empty Map: Time/cost Assessment for Technical Services (Nancy Slight-Givney & Mary Greci)

Givney & Greci took their units through a time/cost assessment to benchmark costs for serials operations. U of Oregon goes through this every five years. For cataloging, the assessment was hindered because the categories were defined for monographs--cessations & other recataloging

modifications were not counted as separate from record correction, since the situation rarely occurs for monographs (outside multiparts).

8. Hot Topics: Serials Consolidation (Keith Courtney-Taylor & Francis, Michael Markwith-TDNet, Kim Maxwell-MIT)

Very lively, very interesting session! The best part was the anecdotes--how these people really worked. Keith Courtney provided fascinating insight into the history of Taylor & Francis--how it survived as an independent publisher, rather than an imprint of Elsevier.

9. Keynote: Wrap-up Panel of Presenters (Jill Emery-U of Houston; Rick Anderson-Univ. of Nevada; Adam Chesler-Ingenta; Joan Conger-U of GA; Ted Fons-III)

[to come]

10. Endeavor Voyager (facilitators Bob Persing-U of Penn; Maggie Rioux-Woods Hole)
Discussion about Unicode & about e-resources.

Unicode: Diane McCutcheon from LC is testing Unicode.

856: Cheryl Fields from NLM would be a good contact on this. Best advice is to create separate holdings records, since this works for display & access. But also leave the 856 fields in the bib records, for searching & maintenance (global change).

III. Announcements:

Next meeting: August 7, for Webmail training.