

Metadata and Meaning: Creating the 21st Century Catalog

4/8/2005

Hogan Center, College of the Holy Cross
Worcester, MA

Morning speaker: Edward T. O'Neill, Consulting Research Scientist, OCLC, Inc.:
[FAST: A Subject Headings Schema for the 21st Century](#)

Breakout sessions:

Stephen Skuce (MIT): *What Happens When Users Create Metadata?*

John Espley (VTLS): *Does FRBR Include Serials?: a FRBR Implementation for All Formats*

Ann Caldwell (Brown University): *Mad about MODS: Implementing MODS in a MARC-Bound Environment*

Boston College staff: *Managing the Metadata Morass: Applying Cataloging Skills Beyond the Traditional Catalog*

Lunchtime speaker: Beata Panagopolous (Harvard University):

[Featured Professional Organization: ASIST, the American Society for Information Science and Technology](#)

Afternoon speaker: Grant Campbell (University of Western Ontario):

[Redefining the Role of Catalogers in the Age of the Semantic Web](#)

Edward T. O'Neill (OCLC, Inc.)

FAST: A Subject Headings Schema for the 21st Century

Ed O'Neill (Consulting Research Scientist, OCLC, Inc.) opened his talk with a brief introduction to the OCLC Office of Research, the entity at OCLC which is undertaking the FAST project. The Office of Research is the largest research organization devoted to Library and Information Science, with ten research scientists, system analysts, and interns. The possibilities for research projects are "almost endless," and the projects themselves are becoming more complex as more library operations enter network space. In conducting its work, the Office focuses on supporting OCLC's membership by maximizing the internal and external impacts of limited resources, making library collections more visible, and developing collaborative projects. Its major research foci are collection and user analysis (working with major research libraries), digital preservation, system and service architecture, knowledge organization and the semantic web, authority control, and metadata schema transformation. Ultimately, there is the intention to support better management decisions through the more intelligent use of data.

FAST stands for Faceted Application of Subject Terminology. The project is a response to the continuing need for subject analysis of documents in the changing information retrieval technology environment. One established response to provision of subject data

in this environment is represented by search engines such as Google, with its thrust toward heavily automated, weighted keyword indexing, without controlled vocabulary. In contrast, twenty-first century library catalogs still largely make use of nineteenth century standards, although those standards are continually updated. The traditional application of controlled vocabularies, although resulting in high quality analysis, is also much more costly than contemporary automated approaches. The FAST project is positioning itself between these two poles, making use of the benefits of traditional cataloging practice, but moving into an automated environment. Other factors driving the project include the great growth in electronic resources and the emergence of numerous metadata schemes, with the latter in particular calling for an approach to subject indexing compatible with the Dublin Core environment. A newly developed subject analysis schema should be simple in structure and syntax and widely usable, provide optimal access points, allow for semantic interoperability, be easy to maintain, and be amenable to authority control and computer manipulation.

FAST has roots in work of the Subcommittee on Metadata and Subject Analysis of the ALCTS Subject Analysis Committee (SAC). One of the major conclusions of the Subcommittee's final report, submitted in 1991, was that it would be preferable to adopt or modify an existing subject schema for indexing electronic resources, rather than beginning "from scratch." (The report is available at <http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/metadataandsujbe/subjectdata.htm>.) FAST is thus completely based on LCSH, the Library of Congress Subject Headings. There are many advantages of working with LCSH. It has rich vocabulary, synonym and homograph control, and is widely translated. It is a *de facto* standard with strong institutional support from the Library of Congress. LCSH has limitations, however, many of them stemming from its development in the card catalog environment. That environment tended to restrict the number of subject headings applied per item, given the need to control the growth of a physical catalog. By their nature, catalog cards also required the use of precoordinated headings.

In an electronic environment, not only does precoordinated syntax require the use of trained and skilled personnel, but in addition, it does not lend itself to automatic indexing or authority control. The great majority of valid LCSH are not separately established, but are constructed according to rules. In OCLC's WorldCat, only two percent of LCSH found on bibliographic records are established in LC's subject authority file. By contrast, 36% of valid, but not established, topical headings appear in multiple records, and 62% appear only once. Being rule-based, these two larger groups are more difficult to control. One example given by Ed O'Neill is the heading: Burns and scalds - Patients - Family relationships. This heading, although more elaborate than most, is nevertheless valid and exemplifies the complex rules for construction which may be invoked. "Burns and scalds" is a heading for a medical condition, and so may take the subdivision Patients. However, "Burns and scalds - Patients" represents a class of persons, which may take the subdivision Family relationships, and so on. The implied rule-based knowledge is clearly difficult to automate.

The FAST Project has several major functional characteristics. As implied by its name, LCSH strings are broken into facets, and all facets are preserved. FAST is format independent, and so may be used with different encoding standards. A FAST heading is

identified by its normalized text, using a modification of the NACO normalization rules. A specific string of normalized text thus becomes functionally equivalent to an authority record number, or ARN.

Authority records are created, in MARC21 Authority Format, for every FAST term in each of the project's eight facets, with the exception, in general, of chronological terms. At present, approximately 1.3 million records have been created. In addition, the establishment of terms is permanent, as authority records for terms will be retained indefinitely in the authority file. If a term (heading or subdivision) needs to be changed, the status of the record will indicate that a change has been made, with information about the change. For example, the heading Trade-unions was changed to Labor unions. The FAST record for Trade-unions will remain, with Labor unions entered in a 750 field, in addition to coding which indicates that this older heading has been replaced with the more recent. This documentation should also aid file maintenance for split headings, as in the case of Nurses and nursing, with was replaced with both Nurses and Nursing.

FAST's eight facets are topical, geographic, form/genre, chronological, personal names (names used as subjects only), corporate names, conference/meeting names, and uniform titles. Files for the first six facets have been developed; the latter two are now under development. *Topical* terms may be established with topical subdivisions if needed. *Geographic* terms have been the most complex to establish, and the following rules have been developed. All geographic names will be established and applied in indirect order, similar to the way in which geographic subdivisions are generally presented. FAST differs from Library of Congress practice in that geographic features which cross national geographic boundaries, such as the Great Dividing Range crossing two states in Australia, will be entered as subdivisions under the smallest first level name in which the place is fully contained. Heading/subdivision combinations will be limited to two levels, except for city sections, which may require three. Qualifiers are used only to distinguish between names, not as aids to searching. Authority records for geographic terms are augmented by information from the Geographic Names Information System (GNIS) database (<http://geonames.usgs.gov/redirect.html>), both as helps for searching and deduplication of identical names. Geographic Area Codes are also included in authority records for every geographic name.

Regarding *form/genre* terms, no form subdivisions will be established under topical terms, but subdivided form terms may be established if needed. *Personal* and *corporate names* are established in FAST if they have both been established in the NACO file and used at least once as subject headings in OCLC. *Chronological* terms, or period subdivisions, may be either single dates or date ranges. Additional text may be used in some headings, as aids to formatting for display, but no other topical information will be included in these terms. Authority records will only be established for periods when the terms are needed as references. Events defined by chronological periods may receive two forms of treatment, with a both a chronological subdivision and a topical term proper for the event represented by the time period.

Among the potential advantages of using FAST is the simplicity of its syntax. A sophisticated understanding of the LCSH system for constructing headings is not required in order to assign appropriate terms in different facets. There is, as well, the potential

"down side" of a loss of specificity in retrieval, resulting from the loss of semantic richness embedded in precoordinated headings. Authority control should also be simplified, as all terms except for most chronologicals are established. Although the potential number of valid FAST headings is great (including those cases in which subdivisions would be established), the actual number in use is much smaller. For example, in LCSH, while potentially over a trillion valid music headings that could be created, few have actually been established.

One interesting outcome of establishing every term is that it is therefore possible for authority files to actually serve as indexes to different facets. Mr. O'Neill demonstrated this possibility through a search on a type of geographic feature, lakes in the state of Michigan. From there, the search was directed toward headings for names of specific lakes, directed then toward specific bibliographic records via a FirstSearch-like interface. Another possible application involves the use of Dewey Decimal Classification (DDC) browser. This graphical, hierarchical interface, with exploding DDC ranges, may lead to headings used in bibliographic records via searching on upper-level Dewey categories.

FAST is available as an OCLC SiteSearch database at fast.oclc.org. The current version is being applied and evaluated in a few environments, and SAC has established a subcommittee to examine the project as well. Future plans include developing the remaining facets, expanding geographic names based on usage data, updating and resynchronizing all FAST headings with LCSH, and revising and expanding the form/genre facet.

Several of the audience questions focused on the relationship between FAST and LCSH. If the FAST project has the need for a new heading, not established in LCSH, it will not be established separately in FAST, but will be recommended to LC for establishment. FAST terms may either coexist with or replace LCSH in a record as desired. It would be possible to have purely local FAST-style terms, but they would not be supported by FAST authority records. Finally, Ed O'Neill emphasized that the project is in no way aiming to discourage anyone from continuing to use LCSH if they have the resources to do so. FAST is not considered as a replacement for LCSH, but a means of adapting and exploiting the richness of its vocabulary in a wider variety of settings.

Reported by
David Miller
Curry College
NETSL Writer/Editor

What Happens When Users Create Metadata?

Stephen Skuce (Massachusetts Institute of Technology)

Stephen Skuce, Rare Books Cataloging Librarian at MIT, led the breakout session called "What Happens When Users Create Metadata?" The presentation focused on the user-created metadata being used in DSpace. DSpace is a digital repository that allows users

to upload their materials and create their own metadata. DSpace is an initiative created by MIT and Hewlett-Packard. It is designed to keep digital materials on the Web with open access for long term preservation. Digital materials submitted may include text documents, images, audio, and video files.

Since users are creating their own metadata, there is less control and problems may arise. Skuce described the results of the user-created metadata as "the good, the bad, and the ugly." "Good" results include receiving complete and rich information such as many key words describing the item, items with author-created abstracts, and the inclusion of series information. The "bad" results can be as simple as a few spelling mistakes and as critical as an author's name appearing in multiple forms, since there is no authority control. The "ugly" problems are harsh on the eyes and usually include metadata written in all caps, the equivalent of information screaming at us in cyberspace.

The problems that occur may be easily remedied by solutions as simple as using spell-check and offering more training for DSpace users. Skuce noted that it was important to focus on the problems that may affect searching. Quality control could perhaps be instituted by staff members. For some reason, it seems as though authors are less likely to consider the information in DSpace in the same way as that in the library catalog. Skuce said most authors uploading items hate birthdates appearing with their names, and do not like punctuation, despite the fact of it being present with their published works as described in library catalogs. Hopefully, the solutions introduced will help limit these problems and allow users the best available access to the materials available in DSpace.

Reported by Danielle Savin
Student, Simmons College Graduate School of Library and Information Science
Boston, Mass.

Does FRBR Include Serials?: a FRBR Implementation for All Formats
John Espley (VTLS Inc.)

FRBR seems to be on everyone's minds these days in technical services. Many of us have read the [IFLA document](#) and know something about FRBR in theory, but lack a clear conceptualization of how it will work in reality. For New England-area librarians who had missed a demonstration of VTLS's Release 45 of their Virtua ILS at ALA Midwinter 2005 in Boston, the breakout session with John Espley (Director of Product Design & Consulting, VTLS Inc.) was an opportunity to glimpse one vendor's conception of FRBR implementation within the serials environment.

Virtua Demo:

Espley opened with a discussion of key "design considerations" VTLS has considered for its product. One was whether to store records as FRBR records or "FRBRize" them on-the-fly. VTLS decided the former makes more sense in terms of collocation, validation checks, and managing linking relationships within a records family. Another

consideration was whether to have a catalog of "pure" FRBR records or a "mixed" catalog with FRBR and traditional MARC records. VTLs opted for a mixed catalog. Espley noted studies at VTLs and OCLC indicating that only 18% of bibliographic records would benefit from FRBRization; the other 82% constitute single occurrences in the catalog without any relationship to other records. Espley also pointed out that Virtua is sufficiently flexible to allow the option of implementing FRBR or ignoring it if desired. To support cataloging - another design consideration - VTLs has created a suite of tools (not demonstrated). Espley described a "FRBRize button" that converts a regular MARC record to FRBR with a single click. Automatic linking between work-expression-manifestation-level records is also possible, as is copying an entire family of FRBR records (a "FRBR tree") from one catalog to another. Virtua also allows one to "batch FRBRize" an entire catalog or even "unFRBRize" records if necessary.

Of the design considerations addressed, it was likely the question of display that aroused the greatest interest, many attendees curious as to what FRBR records even look like. Espley demonstrated VTLs's proposed solution to this design challenge using as an example Beethoven's Symphony No. 6. On the top half of the screen, Virtua's split screen interface displays the work-expression-manifestation relationships within a family of records as an expandable tree structure indented according to the entity level. Distinctive icons denoting each level provide added clarity. On the bottom half of the screen appears the record corresponding to the point in the tree one is highlighting. Each record below the work level possesses both a control number (field 001) and an 004 linking field (appropriated from MARC 21 Holdings Field 004, Control Number for Related Bibliographic Record). The latter corresponds to the control number (001) of the record at the preceding level. Manifestations thus link to their respective expressions; expressions link to the work.

The Beethoven example provided a basic illustration of VTLs's approach to the display of FRBR relationships in Virtua. But what about formats with more complex bibliographic interrelationships, such as serials? Espley's remark, that in searching OCLC for a serial example for his Virtua demo, he had rejected his initial candidate, *Books in Print*, underscored the relevance of this question. With over 200 records in OCLC containing this title, Espley found this serial too unwieldy for use. Instead, he chose *Atlantic Monthly*, with only eleven bibliographic records in OCLC, comprising print and microform formats and incorporating five title changes between 1857 and 1993. To organize their representations, Espley has drawn upon Frieda Rosenberg's and Diane Hillman's concept of the "super work" (as originally formulated by [Rahmatollah Fattahi](#) in 1997). A super work, Espley explained, is an artificial work tying together a family of related works. (In this regard it is similar to a uniform title.) In his demo, he selected "Atlantic monthly super work" at the top of the tree structure; on the bottom half of the screen appeared its corresponding record with the note, "A virtual record for the 'family' of works for Atlantic Monthly." Espley called the five varying titles displayed under this super work "sub-works." At the sub-work level appear the "continues" or "continued by" notes linking one title to its preceding or successive incarnation. Expanding a sub-work level node displays the expression-level node, "Language material-English," providing in the corresponding record below frequency information for that sub-work. Expanding the expression-level node displays the manifestation-level nodes corresponding to the print and microform formats. Their records provide manifestation-level-specific information

including imprint, physical description, reproduction notes, ISSNs, etc. Fully expanding all the nodes on the tree displays in reverse chronological order all the works, expressions, and manifestations under the *Atlantic Monthly* super work, including the eleven manifestations derived from the OCLC records.

Hyperlinking to a related family of works is also possible. Espley displayed the manifestation-level record for "Atlantic monthly (Boston : Mass. : 1857)," which indicated the periodical had absorbed two other periodicals: "Galaxy (New York, N.Y. : 1866)" and "Putnam's magazine." Both display in Virtua as hyperlinks. Selecting "Putnam's magazine," connects one to the "Putnam's magazine super work." Selecting "Galaxy (New York, N.Y. : 1866)," however, brings up an ordinary MARC record for this title. Because Galaxy occurs uniquely as a work, FRBRization is unnecessary.

Discussion and Conclusion:

The questions posed during and after the Virtua demo were varied and generated a lively discussion about FRBR in general and its relationship to serials. Espley noted his own concerns. He believes, for example, that serial catalogers may have overused linking entry fields (especially Field 787, Nonspecific Relationship Entry), which can create unwieldy super works families in the FRBR environment, as was the case with *Books in Print*. Or, as one attendee asked: how can you distinguish between "family" links vs. "neighbor" links? Espley also emphasized the need for "clearer, fuller definitions" of works, expressions, and manifestations. The problem of serial title changes underscores this need: does a new form of title equal a new work or a new expression? VTL's solution, he reiterated, invokes the concept of the "super work." On this topic, he would also like to see clarified the principles for constructing super work titles. MARC tag mappings for FRBRizing MARC records is another concern of his. Espley remarked that, although he had abandoned his own initial mappings in favor of those by Tom Delsey, he suggested some of these may require reassignment. Related to this is his intriguing idea of treating work- and expression-level records as hybrid authority/bibliographic records. He noted that MARBI would thus have to rethink the concept of authority record by allowing them to contain subject headings fields. Throughout today's session, Espley reiterated the necessity of rules and guidelines to provide solutions to these and other questions by providing codification.

The audience had questions and comments as well. In response to the question whether FRBR applied to journal indexing, Espley provided an example of an analyzed issue of *Brigham Young University Studies*. Beneath the work-level record for the journal itself, the issue (v. 35, no. 1 1995) appears in the tree structure at the expression level (enumeration/chronology); the individual articles appear at the manifestation level (author, title of article, pagination, subject analysis (in LCSH), etc.). (One might alternatively consider these articles, or the issue collectively, to be component works within a larger work; this is a matter for debate and reinforces Espley's call for guidelines.) Similarly, someone asked if FRBR could accommodate monographic serial analytics, to which he responded affirmatively, but provided a music analytics example: a single manifestation-level record (a sound recording) linking to separate work-expression-level records for three compositions by Mozart. Though interesting, this was really a different situation; in the case of monographic serial analytics, as with journal

indexing analytics above, does one treat analytics as component works within a work, or as separate expressions within a work?

Other concerns expressed included the responsibility for record cleanup and the implications for shared cataloging and bibliographic utilities. Espley noted that, although, the records he had taken from OCLC for the *Atlantic Monthly* example had not required clean up prior to FRBRization, another set of records which he had FRBRized had. As for shared cataloging, he remarked that the utilities must still address the challenge FRBR poses. In the meantime, unFRBRization is the current solution; he mentioned a university library in Belgium that does this prior to sending their records to a union catalog. In addition to this library, he noted a public library in Virginia that has already adopted FRBR.

Espley's response to his own question, "Does FRBR Include Serials?," is "I think it can." Despite the problems addressed today, FRBR, he believes, will improve OPAC displays and help to realize the Paris Principle of collocation. Again, he awaits more rules to refine and guide practice. Time constraints prevented more discussion. It would have been interesting, for example, to observe how FRBR (and Virtua) handle holdings and item records. A comparison of how different ILS systems handle the same serial title in FRBR would also be useful. One hopes that in converting the more theoretical constructs of the IFLA document into the more pragmatic codification represented by AACR3 (now *RDA: Resource Description and Access*), the Joint Steering Committee for Revision of Anglo-American Cataloguing Rules will derive inspiration from some of the more practical solutions offered by ILS vendors, such as VTLS.

International Federation of Library Associations and Institutions. *Functional Requirements for Bibliographic Records: Final Report* (München: Sauer, 1998). <http://www.ifla.org/VII/s13/frbr/frbr.pdf> (accessed April 27, 2005).

Frieda Rosenberg and Diane Hillman, "An Approach to Serials with FRBR in Mind: CONSER Task Force on Universal Holdings," (draft document, last rev. 1/24/04):1, http://www.lib.unc.edu/cat/mfh/serials_approach_frbr.pdf (accessed April 27, 2005).

Reported by Craig K. Thomas
Cataloger, Germanic Division
Harvard College Library Technical Services

Mad about MODS: Implementing MODS in a MARC-bound Environment
Ann Caldwell (Brown University)

Using her institution's experience as an example, Ann Caldwell (Metadata Specialist, Center for Digital Initiatives, Brown University) provided NETSL attendees with a picture of what MODS is and how it is applied in digital libraries.

The MODS Schema:
MODS (Metadata Object Description Schema) is a bibliographic element set developed by

the Network Development and MARC Standards Office at the Library of Congress particularly for library applications. The need for a set of descriptive metadata in XML that is simpler than MARC but rich enough for complex digital objects was the stimulus that led to the development and adoption of MODS. Designed with hierarchical structure and compatibility, MODS satisfies the prospect for rich description and reduces the constraints of cross mapping between different metadata schemas. Using language-based tags rather than numeric ones, all MODS elements can be expressed in XML and are semantically parallel to MARC. There are no rules (e.g. AACR2 or ISBD) laid down for the content of bibliographic description marked up by the structure, whereby it is easy for non-professional catalogers or even laypersons to learn. The MODS element set includes title, name, type of resource, genre, language, physical description, abstract, table of contents, target audience, note, subject, classification, related item, identifier, location, access conditions, extension, and record information. All elements are repeatable but not required. There are two types of elements in MODS: simple elements, such as

```
<note>March for voice and piano.</note>
```

and wrapper elements, such as:

```
<titleInfo>
```

```
<nonSort>The</nonSort>
```

```
<title>service flag</title>
```

```
<subTitle>a song</subTitle>
```

```
</titleInfo>
```

Information about MODS is available at <http://www.loc.gov/standards/mods/>.

The Center for Digital Initiatives at Brown University:

Formed in 2001 as the production arm of the library's Digital Services Department, the Center for Digital Initiatives serves the entire university for its digital undertakings. The effective usability of name element, publication information and related item elements convinced the Center to adopt MODS as the descriptive metadata schema for cataloging digital objects. MODS records are either converted from MARC records or created from scratch. Staff with cataloging background, graduate students, undergraduate students, library school students and faculty all participate in the process of MODS record creation. Tools for MARC-MODS conversion and creation are available. Some, such as NoteTab, are as inexpensive as \$20. Although no element is mandatory in the MODS standard, eight elements - title, name, origin information, language, physical description, note, related item and identifier - are required for each MODS record to be entered into the database. While LCSH is not assigned as a subject heading schema, various controlled vocabularies and thesauri are used to ensure consistency, improve search results, and enable record sharing. Personal names linked to MODS records by ID numbers are maintained in an authority database. Collections existing in the Center are Carriers' Addresses, African-American Sheet Music, Alcohol, Temperance & Prohibition and World War I Sheet Music, etc. Information about the Brown University Center for Digital Initiatives is available at <http://dl.lib.brown.edu>.

Reported by Mary Wu
Roger Williams University

Managing the Metadata Morass: Applying Cataloging Skills Beyond the Traditional Catalog

Group presentation by Boston College staff:

Chris Conroy, Associate University Librarian for Collection Services

Betsy McKelvey, Electronic Resources and Technical Services Librarian

Este Paskausky, Digital Library Systems Analyst

Meg Critch, Digital Resources Cataloger

This group presentation, by both library and non-library Boston College staff, painted a compelling picture of both the benefits to technical services librarians of being involved with institution-wide digital projects and the importance to the institution of involving professionals with expertise in metadata creation and the implications for the public of database structure and development.

Chris Conroy began the discussion by stating that the cataloging department at Boston College is reinventing itself to remain a viable and important player in the digital area. There have been significant shifts in the library budget in the direction of electronic resources, resulting in a discrepancy between budget allocations and the percentages of employee time devoted to different types of materials. Another important factor in the transition at BC is the upsurge in available digital asset management tools, especially for archival or special collections. There is an emphasis on working together as a team, because an environment dominated by changing or uncertain standards makes teamwork even more important than usual. They are "thinking beyond the OPAC" in that the emphasis is no longer on a single, integrated access/retrieval tool, but rather a series of linked tools. These include DigiTool, an ExLibris digital asset management product; an Electronic Resources Management (ERM) database developed in-house; MetaLib for federated searching; SFX for reference linking; the Aleph OPAC/ILS; and Digital Commons, a ProQuest product for institutional repository management.

Ms. Conroy addressed the critical question of how a staff makes time for such a range of new responsibilities, while not abandoning the traditional ones. The approach at BC is to switch to a "project-oriented focus," which treats everything (including traditional cataloging) as a project. This mindset, they find, helps to manage priorities and time better than trying to wedge new responsibilities into a menu of fixed duties. In addition, the department has undergone an "attitude adjustment": when asked to take on a new responsibility, don't say no, but rather figure out how to do it. It is essential that cataloging management gets a seat at the table when digital institutional projects are discussed. Not saying "no" allows this to happen, along with building and maintaining partnerships across the institution. Catalogers at BC are no longer defined only by ownership of traditional tasks, but are additionally known as consultants and partners, offering their expertise to other projects.

Betsy McKelvey discussed the development of BC's ERM database. She began working on this project in September 2003, and at present does not work in cataloging at all. The ERM resource was developed in-house, and is currently used only by library staff in collection management. It functions as an information hub for MetaLib, SFX, the proxy

server and Serials Solutions profile, and is linked to the Aleph acquisitions module. It may be asked, is it necessary to have an ERM database? Would good use of the Aleph system alone be sufficient to meet the requisite information needs? Examination of the Digital Library Federation's (DLF) ERM Initiative specifications demonstrated that, in fact, this would not be the case: the ILS alone would not be able to handle the job. Development of the ERM was also important in terms of broadening staff participation in the project, given the Boston College Library's "learning organization" philosophy. Finally, the database was developed in-house, as the staff did not wish to wait for an adequate vendor product.

The ERM database was prototyped using Microsoft Access to test the DLF entities needed and to develop metadata crosswalks from subscription agent information. A few elements not in the DLF specifications were added to the database. The Access database allowed for testing and interface design, and was used in parallel with the Web-based ERM while the latter was being developed. Cataloging expertise was brought to bear on a systematic approach to creating metadata in the ERM. The same quality standards, using AACR2 and the LCRIs, are applied as in mainstream cataloging. This has proved to be a way of saving time as well as achieving consistency, largely because those creating ERM metadata are already familiar with these standards and don't need to ask questions about them.

Este Paskausky discussed BC's developing use of DigiTool, a Digital Asset Management system (DAM) product from ExLibris. DigiTool "creates a space for digital objects that would be in library collections if they were physical." Management of these objects involves technical and management as well as bibliographic metadata. The DAM provides a framework for maintaining these types of metadata, linked by an Oracle database. It also collates the metadata with the objects themselves, which the traditional ILS does not do.

DigiTool has a public, web-based front end, but the present focus of the project is on cataloging issues, including data cleanup, to ensure record portability. The BC cataloging department has played an instrumental role from the beginning of this project, which was driven initially by the digitization of the slide library collection. "Systems people" are successfully collaborating with catalogers, bringing together both disciplines' approaches and concerns. An upcoming phase of this project will be the migration of special collections data from the Aleph catalog to DigiTool. The challenge here is that archives metadata is very different, so there will be the need for interoperability. What is unique about DigiTool is that it is possible for various metadata standards to be integrated, addressing this need. BC is also looking toward the implementation of federated searching. Laying the basis of shared standards now will allow DigiTool records to be searchable when that becomes operable. Additional anticipated future stages include the development of preservation metadata, which again will benefit from the work of catalogers; incorporation of EAD (Encoded Archival Description) finding aids; page-turning and METS support; and possible incorporation of the institutional repository (IR), presently a separate system.

Meg Critch concluded by discussing the IR, which has pilot project status at this point. BC is using Digital Commons, a ProQuest turnkey system product, for the project. The IR

was initially populated with electronic dissertations from 1997 to the present; current discussions focus on the types of materials that might be included (offprints, datasets, working papers, etc.), and determining who else at the university is interested in the pilot project. Again, catalogers have played an important role. At the beginning of the IR project, they asked the questions which nobody else was asking, including authority control and planning ahead for how users will search the database. Catalogers were also aware that some items going into the IR were already cataloged fully (as tangible objects) in the ILS, and so established the need to build the link between the two systems. In short, having catalogers involved with digital projects points to how it is "essential to have the right people asking the pertinent questions at the start."

Reported by David Miller
Curry College
NETSL Writer/Editor

Beata Panagopoulos (Kennedy School of Government, Harvard University)
American Society for Information Science & Technology (ASIST)

(Following is the text of Ms. Panagopoulos's talk.)

I'm very happy that I have this opportunity to talk to you all about the professional association that I'm active in, the American Society for Information Science & Technology, otherwise known as ASIS, ASIST or A.S.I.S. and T. I'm Head of Technical Services at the Kennedy School of Government Library at Harvard and I've been in cataloging and technical services for over 23 years. So, I feel particularly qualified to tell you why you should consider joining ASIST and what you would get out of it.

I joined ASIST after attending a conference here at Holy Cross in the fall of 1991. The program was called: "Standards, Applications, and Document Delivery: Why Should I Care?" Cliff Lynch was the keynote speaker and I was really impressed by how fascinating and important the topic of standards development was. Little did I know that there aren't that many people apart from Clifford Lynch who can make standards sound exciting, but his talk did hit a nerve: it placed my work as a cataloger in the much broader context of international scholarly communication and exchange, and increased my appreciation of how difficult international standards are, yet how critically important they are for the free flow of information. So when in 1993 I was offered the opportunity to work on the program committee for the New England Chapter's conference on digital imaging, I couldn't refuse. That program, "Image/Imaging: a Showcase of Digital Imaging Applications" was spectacular. It featured a pre-workshop with internationally renowned image and multimedia database expert Howard Besser, followed by a full-day workshop with five breakout sessions. This local program was just as professional and interesting as any conference I had attended nationally.

In those two examples of ASIST programs, you have a microcosm of what I love about the organization. Both conferences featured internationally respected experts in information science and technology - the one talking about standards, which is directly

related to my work, and the other about image databases, which, although they are integral to library collections and catalogs today, had little to do at the time with my day-to-day responsibilities. ASIST brings together people from all over the information studies spectrum, to share their research and practice. It is very rare that you walk away from an ASIST program without having that "ah-ha!" reaction - as you make connections with your own interests.

Although I feel proud that these programs took place in 1991 and 1993, just as the World Wide Web was emerging, ASIST was actually founded way back in 1937 as the American Documentation Institute. Its initial interest was in the development of microfilm as an aid to learning: developing microfilm readers and cameras; promoting negotiations and research in the area of photo duplication of copyrighted materials; and support of Interlingua, an early rival of Esperanto to foster international science communications. In 1968, the organization changed its name to the American Society for Information Science, reflecting the fact that the membership, which had increased sevenfold, was concerned with all aspects of designing, managing, and using information systems and technology. Today, as information and communication technology are part of our daily lives, ASIST members come together to examine all aspects of online databases (from their technical structure to their social consequences), their use in government, industry and education, and the development of the Internet and the World Wide Web. In 2000 the Society changed its name to the American Society for Information Science & Technology to reflect the range of interests of its members.

So, who are ASIST members? ASIST is a very diverse association. It's composed of librarians, information architects, knowledge managers, vendor representatives, corporate information specialists, and university faculty, to name just a few groups. The membership is commonly thought of as split between "practitioners" and "academics." There are, in fact, more practitioners than academics, even though ASIST's publications are heavily oriented towards academic research and theory. In order to address this split, the next annual conference, which will take place in November in Charlotte, North Carolina, is called "Sparkling Synergies: Bringing Research and Practice Together."

As technical services librarians we're interested in access to information. Many presentations at ASIST conferences focus on access, information retrieval, database design, and information architecture. Whether you're interested in studies on information seeking behavior or a panel discussion of "the nature of a work", digital libraries and metadata projects or scholarly communication and electronic publishing, ASIST brings it altogether with an interdisciplinary approach. No, you won't find panels about AACR3, but you will find discussion of the Dublin Core and FRBR. Each and every one of us has experienced radical change in our jobs. And being involved in ASIST can help you gain the skills and knowledge you need to advance your career.

What I love in particular about ASIST is its structure. You can be a member of the national organization and get a lot out of going to the annual meetings and reading its publications. In the very near future, membership will also give you access to the ASIST Digital Library. But like many associations, there are also special interest groups and chapters you can join. Here are a few of the SIGs which may be of interest to you: Library Technologies, Digital Libraries, Classification Research, Visualization, Images and

Sound, and Knowledge Management. The SIGs, apart from bringing together people with like interests, are also the backbone of the annual conference. The SIGs put together most of the panel discussions. So, an ASIST conference is really a grassroots production. There isn't a corporate firm putting on the meeting - it's the members who are doing it, which gives the conference a lot more meaning.

Then there are the chapters. The New England Chapter is one of ASIST's most active and successful. I highly recommend becoming involved at the local level for a number of reasons. You have the opportunity to learn leadership skills. You'll learn how to plan and produce great programs, you'll have the opportunity to do public speaking, and you'll meet some really interesting and inspiring people. It's very easy to become involved with the New England Chapter - there is no bureaucracy to cut through. You just say you want to participate and they'll tell you when the next meeting is. This past December the Program Committee, very ably chaired by Beatrice Pulliam, put on a great program at MIT called "Freedom vs. Control: Rights Management in the Digital Age". And on May 3, they are presenting another program at Providence College called "Syndicate, Aggregate, Communicate: New Web Tools in Real Applications for Libraries, Companies and Regular Folk." The program will be about blogs, wikis, RSS, Instant Messaging, Chat, and Folksonomies. It will address the questions: How and when do these tools work together? How can you use them in your environment? How do you convince your boss that they are worth implementing?

Well I could go on and on about ASIST, but I think you've heard enough for now. Does anyone have any questions?

Grant Campbell, Assistant Professor, Faculty of Information & Media Studies, University of Western Ontario:

Redefining the Role of Catalogers in the Age of the Semantic Web

Grant Campbell's stimulating talk not only served to introduce the basic concepts of the Semantic Web to those perhaps unfamiliar with them, but also stimulated a lengthy question-and-answer period. The concept of the "Semantic Web" is still somewhat new. It is the brainchild of Tim Berners-Lee, originator of the World Wide Web. The development of the Semantic Web is a collaborative effort led by the World Wide Web Consortium (<http://www.w3.org/2001/sw/>). But what does this phrase mean? What is the Semantic Web? At its core, the Semantic Web is built on machine-understandable data, as compared with machine-readable data. The latter, with which librarians are very familiar, is electronically retrieved, sorted and categorized. In the Semantic Web, accurate and appropriate metadata should also be machine-understandable, meaning that the user may pose an intelligible question, and retrieve answers which assume contextual understanding of the question's components. This will stand in strong contrast to the current dominance of retrievals based on text ranking.

The development of the Semantic Web depends on three basic elements: XML (eXtensible Markup Language), RDF (the Resource Description Framework), and OWL (Web Ontology Language). XML is derived from SGML (Standard G Markup Language), a

standard which has proved to be too complicated to find common use. By contrast, HTML, the other well-known derivation of SGML, has proved to be unsatisfactory in that it only describes the appearance of online documents. XML is simpler than SGML, yet allows the data found in documents to be encoded according to its semantic meaning. The broad implementation of XML will allow access to the "deep web" of information available online but existing in formats not generally accessible to today's search engines. As an example, the data in an Access database may be exported in XML, thereby making it accessible to search engines. Another important consequence of XML implementation is likely to be a dramatic increase in "on-the-fly" web pages, whose content is generated according to an "occasion", such as a particular query. In an era of multiple, overlapping markup standards, XML serves the important function of a switching language, a means of converting data from one form to another.

RDF is "an XML-based model for making statements about resources." At its core is a simple syntactic structure: Subject - Predicate - Object. This structure, consisting of a "triple," allows statements to be made about objects. Its typical formulation would be "Resource A - has an attribute - with value B," and a concrete example would be "The movie Solaris - was created by - Steven Soderbergh." The following assumptions lie behind RDF:

1. It can be expressed in XML but is not dependent on XML. RDF is a model, similar to AACR2, not a specification such as MARC21.
 2. Everything can be assigned a Uniform Resource Identifier (URI). URIs may be assigned to people, emotions, ideas, and so on, as well as web sites. There can be many different sources of information found at URIs: for the example above, both the movie Solaris and Steven Soderbergh have individual pages available via the Internet Movie Database.
 3. Information is best organized from the ground up, not the top down. The best form of control is that which approaches no control at all. People should be allowed to organize their own connections between resources and concepts.
 4. Information organization is a matter of making statements about resources which both have meaningful contexts and make sense within an information system.
- In the case of the movie Solaris, XML-encoded RDF statements about it could serve as the basis for sending an agent out on the Semantic Web to find statements about the movie itself, the director Steven Soderbergh, and the concept "filmmaker."

The third important element, "ontology," is defined as "a machine-readable expression of a shared conceptual framework." An ontology is usually expressed as a domain-specific combination of a classification scheme and controlled vocabulary. Ontologies are designed to link similar concepts in different namespaces, e.g., the concept of "director" in the film and theater domains. They are designed to increase a search agent's ability to find and relate similar information in different domains.

These elements may be put together to envision the Semantic Web in a layered model. At the base are Unicode and URIs. Above that comes data expressed in XML, with RDF statements about that data. Next comes the layer of ontologies which link the RDF statements together, creating the context for logical inferences that can be made by computers. At the very top are digital signatures, guaranteeing the authenticity of the data, and finally, the concept of trust.

At present, libraries and library catalogers use the World Wide Web "as a huge irrigation channel" for swiftly and efficiently interchanging records. For example, simplified pointers to fuller catalog records are exported in Dublin Core for transfer to Open Archives Protocol sites. However, it may be possible "to make things a bit more fun."

Mr. Campbell offered some thoughts about "repositioning" the traditional strengths of library information organization in the era of the Semantic Web. The first question to ask would be, "What's it like beyond our walls?" The world is full of experts and enthusiasts. Their work is valuable and shared, and we should make use of it. Human culture is saturated with complex, nuanced, important relationships. We cannot represent these in any kind of profound way via traditional schemas. In addition, scholarship has become increasingly interdisciplinary, and popular culture has become more of a subject of scholarly study. All of these factors point to the need and the opportunity to bring our distinct strengths to the job of providing meaningful contexts for the knowledge bases that already exist. What librarians can bring to the table are our skills in information resource description, organization and access; information evaluation; and a rich, sophisticated body of theory on bibliographic relationships, of which few others are aware.

What might bibliographic description be like under these circumstances? Using a Dickens novel as an example, there is first the level of item-specific information, such as the location of a single copy. This data can only be provided locally. Different pieces of bibliographic information, such as that about the author, the work, and related works, may be separately broken out. RDF statements may be used to reference bibliographic information from the Web for each of these elements, including types of descriptive and analytical bibliographical information not included in library descriptive cataloging. In searching for reliable sources, such as standard bio/bibliographies for author information or information about the social contexts of works, and expressing them in RDF statements, catalogers may take on roles similar to those of collection development or reference librarians.

A second "repositioning" question would be, "What we can do for others on the Semantic Web?" Human minds are full of ideas, and the arising of ideas largely depends on the existence of documents. The intensity of an idea in one or many minds is based on encountering a document repeatedly, in a variety of contexts. What we do in libraries is make it possible to "bump into" texts repeatedly, providing the possibility of those multiple encounters. We are committed to the accretion of significance taking place in this way, and this commitment will not become obsolete in the foreseeable future. Through our work in the Semantic Web, we may help to define bibliographic universes in ways meaningful to those using them.

At the outset of his talk, Grant Campbell emphasized his conviction that the theory of cataloging has beauty and depth which does not exist in most disciplines. This theory, which is far too little known, can easily inform our work in the Semantic Web. As is so well articulated by the late Seymour Lubetzky, catalogs exist importantly to bring together all the works by a particular author, and all the editions/expressions of a particular work. Works are not simply discrete entities, but are entities that have relationships important to different people in different ways. Within the Semantic Web,

we may provide not only a group of RDF statements, but an intellectual structure which makes these statements meaningful.

Topics raised by audience questions were wide-ranging, from the potential redesign of the catalog, to mapping between languages and cultures and the migration of MARC data to XML as an essential starting point. Questions were also raised regarding the creation of RDF data, sharing RDF-based records, and the permanence of sources from which RDF data is drawn. One audience member asked if the rise of the Semantic Web would mark "the death of the OPAC." This will not necessarily be the case, as the idea of the catalog will likely be strengthened. The tension between the catalog as inventory and as a research tool will continue to be negotiated, but surrogate descriptions will become more important, not less, as time goes on.

Reported by
David Miller
Curry College
NETSL Writer/Editor