Notes from Some Meetings of Encoding-Standards Interest at ALA Midwinter 2017

Report prepared by Jim Soe Nyun, Chair, MLA Encoding Standards Subcommittee

January 29, 2017

MARC Advisory Committee
Saturday, January 21, 8:30-10:00
Sunday, January 22, 3:00-5:30

Proposal No. 2017-01: Redefining Subfield $4 to Encompass URIs for Relationships in the MARC 21 Authority and Bibliographic Formats

NOTES: It was acknowledged that it would be complicating things to redefine $4 across the Bibliographic and Authority formats to both a) conflate “relator” and “relationship” codes and b) permit URIs to be entered in the subfield. Systems that currently display the subfield or act on the value would need to be modified. However, the authors of the paper made a persuasive case that it was essential to be able to record the URI for the relationship expressed in $4. It was pointed out that adjacency of the URI to its corresponding value—if present—didn’t really need to be a concern; the term in $4 and the URI were different systems of naming the relationships and didn’t need to be synced in the field; users would be free to supply terms only, URIs only, or both. Proposal passed.

Discussion Paper No. 2017-DP01: Use of Subfields $0 and $1 to Capture Uniform Resource Identifiers (URIs) in the MARC 21 Formats

NOTES: Paper introduced by Stephen Folsom, who spoke of the need to differentiate URIs that represent the thing itself (real-world objects, RWOs) from those that represent authorities or pages. This is a distinction of much importance in linked data. Defining $1 for RWOs would take up the last-available subfield across the format. Some discussion that this would be a worthy use of the final subfield. The paper will return as a formal proposal asking for $1 to be defined for URIs that will dereference to the RWO.

Discussion Paper No. 2017-DP02: Defining Field 758 (Related Work Identifier) in the MARC 21 Authority and Bibliographic Formats

NOTES: Paper introduced by Chew Chiat Naun, who indicated that he wanted to keep the proposal neutral in its stance towards the FRBR Work. Some discussion that “related work” was not an entirely correct name for the proposed field. Other discussion that it would be desirable to develop an indicator indicating whether a field is for a work or expression, plus an option for “no information,” as well as another for N/A in the Authority format. Will return as a proposal incorporating these suggestions. The paper already incorporates the use of $4 for relationship URIs discussed in Proposal 2017-01.

Proposal No. 2017-02: Defining New Subfields $i, $3, and $4 in Field 370 of the MARC 21 Bibliographic and Authority Formats

NOTES: Presented by Adam Schiff. Approved provisionally with the proviso that some wording
identified by reviewers in advance of the MAC meeting would be clarified.

Proposal No. 2017-03: Defining New Subfields $i$ and $4$ in Field 386 of the MARC 21 Bibliographic and Authority Formats

NOTES: Passed with some suggested rewording to clarify the definition of the subfields. We will need best practices on how to apply the newly defined subfields.

Proposal No. 2017-04: Using a Classification Record Control Number as a Link in the MARC 21 Bibliographic and Authority Formats

NOTES: There was some concern that the classification used in a bib or authority record might not corresponding exactly to what would appear in a classification record. This would be a concern in the LC shelflist where tables could be used to extend the classification proper. The paper was submitted by the German National Library whose classification numbers do not have this problem. Proposal passes. Users will need to be careful that the $a$ of their record corresponds to any classification record cited.

Proposal No. 2017-05: Defining a New Subfield in Field 340 to Record Color Content in the MARC 21 Bibliographic Format

NOTES: MLA submitted a suggested rewording of the first sentence of the proposed subfield, from “Presence of color in the content of a resource” to “Color characteristic of the content of a resource.” Rewording accepted, proposal passed unanimously.

Proposal No. 2017-06: Adding Subfields $b$, $2$, and $0$ to Field 567 in the MARC 21 Bibliographic Format

NOTES: Matthew Wise present this and the following proposal on behalf of the National Library of Finland. This paper defines a way to describe that a work is an example of a method of inquiry rather than being about that method. An example given was that of a book that employed Shenkerian Analysis to reach its conclusions, but itself was not about Shenkerian Analysis. Proposal passed.

Proposal No. 2017-07: Adding Value “No information provided” to the First Indicator of Field 070 in the MARC 21 Bibliographic Format

NOTES: The National Library of Finland uses NAL-style call numbers for some content and wants a way to not have to determine whether a resource is in the NAL collection. Defining value “[blank]” in the first indicator position would let them use the NAL scheme and be agnostic about coding whether the item is owned by the NAL. Some opposition to this change, in part from worries about how this would impact NAL and their legacy data. Still, item passed, with 3 opposed, 2 abstaining.

Discussion Paper No. 2017-DP03: Defining New Fields to Record Accessibility Content in the MARC 21 Bibliographic Format

NOTES: The authors withdrew from their paper the strategy of repurposing of the second position of the 007. There was some concern about what this paper might mean for resources cataloged according to the provider-neutral model; in theory different accessibility presentations
would constitute differences in expression that would justify a new record, but there still could be differences in how a resource is presented. Overall much interest in developing this into a formal proposal.

This paper steps into a confluence of at least three groups that are dealing with this issue: CCM, OLAC and the ALCTS/LITA Metadata Standards Committee. The CCM discussion paper is the most developed at this stage. OLAC is very interested in collaborating on a proposal to be developed out of this discussion paper since many of their resources have accessibility issues; also, there was a working group set up in OLAC to deal with accessibility metadata. The Metadata Standards Committee may have something to contribute so I offered to share contact information. There are still a lot of questions about what information needs to be recorded that the final proposal will attempt to research the topic further.

Discussion Paper No. 2017-DP04: Defining Subfields $u, $r and $z in Field 777 of the MARC 21 Bibliographic Format

NOTES: Generally not controversial. However, question 5.6 of the paper asked whether we could expand these subfield definitions for Fields 760 and 762 without specific use cases. The Committee would prefer seeing use cases and not create an automatic expansion of the format. Will return as a proposal.

Discussion Paper No. 2017-DP05: Providing Institution Level Information by Defining Subfield $5 in the 6XX Fields of the MARC 21 Bibliographic Format

NOTES: This paper makes sense when applied to the closed world of the GND, which initiated the paper. When expanded out to other communities it could make things quite complicated. The $5 is often used to tag local data that some libraries delete automatically, and content that might be useful from GND libraries might get stripped out. Some concern that using $5 for terms from alternate vocabularies might leave people feeling that they might need to duplicate the field with their own $5 if they’d like to use the same term, a possibility for serious record bloat. Definitely not unanimous support, and some outright strong opposition on the Committee. Unresolved as to whether this will return as a formal proposal.

Business meeting/Library of Congress report/Other

NOTES: Some discussion of how to implement the fast-track approval process. Matthew Wise proposed that these be bundled with the MARC update cycle during which they are approved. Also, the changes would appear in red as with changes from formal proposals, only the fast-track changes would not show up in the Content Designator History. And suggested applying the same 60-day embargo against using the changes as other MARC changes are subject to. Agreement from the room.

Three fast-track proposals are either approved or pending. MLA’s request to make Field 384, Key, repeatable has been tentatively approved pending final finessing of some details. This proposal is a development of a withdrawn discussion paper put forward by the Koninklijk Conservatorium Brussel.

A general heads-up: New IFLA LRM concepts are likely headed our way, and MARC may need to
accommodate them.

Online Audiovisual Catalogers, Cataloging and Policy Committee (OLAC CAPC)
1/20/2017, 7:30-9:30 p.m.
Excerpts of topics of special interest to Encoding Standards

MARC Advisory Committee report (Bruce Evans for Cate Gerhart) Most papers and proposals are working toward making data elements more discrete, more likely to make linked data work better. There’s a discussion paper, DP2017-03, that is looking at ways to encode information about accessibility features of a resource. There’s a video accessibility task force within OLAC and this might be an opportunity to collaborate on the final proposal that might come out of the discussion paper. The paper looks to repurpose the no-longer-used 008/02 and possibly establish a 341 element field and a special 532 note. Would like to have access to a richer vocabulary for the fields to employ than currently exist, e.g. one that has closed captions apart from just captions.

OCLC report included a mention that OCLC has partnered with the Internet Archive to ensure the durability of Persistent URLs (PURLs) (Announcement from September 27, 2016: https://blog.archive.org/2016/09/27/persistent-url-service-purl-org-now-run-by-the-internet-archive/)

OCLC Linked Data Roundtable: Stories from the Front
Saturday, January 21 10:30-11:30
#OCLCLDRT

Sally McCallum
"Save my Bibliographic Data"
Working towards second BIBFRAME test, hoping to “plan better” than for first. Converting 19M bibfile records into BF 2.0, to finish in about a week.
MARC conversion issues: changes in descriptive conventions; variety with level of correctness; duplication within the format, some coded versus free text.
Quote of the weekend: "If MARC dies it will be through obesity"

John Chapman
"Transforming Digital Material Metadata into Linked Data"
Work at turning metadata about digital objects into RDF.
OCLC and OCLC Digital Collections Strategic Advisory group, going from CONTENTdm into RDF.
Library metadata is wildly heterogeneous; lossy transformations makes for less useful output; TSV or XML => JSON => Analysis (list field types, assign types, external vocabularies)
Generates preview, then goes to reconciliation then transformation into RDF.
Feedback: Mapping of field type (e.g. into Schema) is not perfect; some incorrect matching; need manual tools from no match; response time issues.
Uses EntityJS to help find entities.
Future: Work with DPLA service hubs; Enhance user experience by building larger graphs; explore tools for md creation, not conversion

Phil Schreur
"Linked Data for Production: a multi-institution approach to technical services transformation"

Intro to the LD4P project

Summary of work

--Documenting change recommendations
--Target ontology development; working on the core BF ontology
--Columbia, intersection of libraries and museums; mapping of current description into BF; not good at first, most data into notes; working on art object ontology extension
--Cornell, two projects. 1. Extension for rare materials. 2. Metadata creation for non-commercial hip hop LPs, including cataloging things in MARC and others in BF, comparing results
--Harvard, cartographic and geospatial datasets; developed use cases; used models to create metadata to check on usefulness
--LC, 4 projects. Archival and film recorded sound collections. Print and photo resources; BF 2.0 development; RDA and BF issues
--Princeton, annotations with Derrida collection
--Stanford. Performed Music Ontology; Tracer Bullets, Blacklight and linked data
--Linked Data for Production wiki available. "Just Google it."

Post-presentation discussion notes:

Jenn Riley has just published "Understanding Metadata" a revision of an early NISO pamphlet (http://www.niso.org/apps/group_public/download.php/17446/understanding%20metadata). Phil Schreur mentioned that they’d be discussing with original catalogers what it would take to begin cataloging natively in BF. Scared of the update issues. What would this mean for PCC libraries?

OCLC working with some Dutch libraries to visualize metadata.

What to do with data that is updated? PS mentioned that they were most interested retaining all information, but just displaying the corrected or updated data.

MARC ==> RDF has been more complex than what people thought it would be. Some items are in MARC that doesn’t make sense in BF. MARC still kept alive for some processing purposes. Issues dealing with material types in MARC, a big issue for OCLC. A model of MARC under the surface with BF/LD on top.

LC has committed to maintaining a full MARC record in their current work.

PS, pet peeve with MARC. It was designed to recreate a catalog card. Blob of information. Need a way to make MARC linkable and participate in the larger worlds.

RT: Work to move from data design to description to discovery.

MARC Format Transition IG
January 21, 3-4 p.m.

[Arrival delayed by Atlanta’s Women’s March along shuttle route and missed first two-thirds of meeting]
Arrived mid-discussion of structured topic strings versus generally single-terms from the FAST vocabulary. Still much sentiment to maintain topic strings and the richness they offer experienced library users. Implementing FAST in linked data is easier because the terms are set up with URIs attached to the FAST fragments derived from LCSH. Complex LCSH strings are seldom set up. But there are many worries that if we were to move to using FAST of LCSH we'd be leaving LCSH for expediency.

Comment about discovery at the entity level with linked data, versus at the record/object level.

Faceted Subject Access Interest Group
Saturday, January 21, 4:30-5:30

Looking for someone interested in being a co-chair for the IG.

John Chapman
Update on OCLC FAST
OCLC Research will be surveying FAST users.
FAST list is available.
CA. 90M records in OCCL have FAST.
Reminders that OCLC has FAST tools, to search and supply FAST headings.
FAST downloads are available.
All of FAST can be downloaded.

Janice Young
"Enhancing Access to Resources with LC’s Faceted Vocabularies"
General intro to LC’s faceted vocabularies.
History: Began in 2007 with LCGFT.
LCMPT, started work in 2009.
LCDGT, development began in 2013.
Some existing terms will be cancelled if they conflate demographic information with another term.
Structure: LCGFT: Single terms all have single highest term. Some terms are combinations of other terms and may no longer accurately live in a single vocabulary.
LCMPT: No broader terms for three top terms in thesaurus
LCDGT: Not a hierarchical thesaurus; very few broader terms
Purpose: Simplify metadata creation; to provide a better discovery experience
Assignment rules: Assign multiple terms to describe multiple aspects
Do not subdivide
Assign subject headings as usual, including subdivisions

PDFs of these vocabularies are available but Class Web is more up to date.

Metadata Interest Group (ALCTS)
Sunday, January 22, 2017 8:30 AM - 10:00 AM
Location: GWCC, B204

Presentation Title: Automating XML remediation with Python’s lxml package and schematron
Presenter: Jeremy Bartczak – Metadata Librarian
Affiliation: University of Virginia
Abstract: The University of Virginia (UVa.) contributes thousands of digitized photographs to the Digital Public Library of America (DPLA). Plans are underway to submit additional objects from multiple legacy digital conversion projects. These projects were implemented in MODS over the course of several years. As local policies evolved, descriptive metadata practices differed across collections. The UVa. Library’s Metadata Analysis and Design team is now in the midst of a large-scale project to remediate this data. Thanks to detailed documentation online about the
DPLA’s metadata application profile, and helpful analysis from DPLA staff, a strategy has been implemented to ensure consistent metadata display for UVa. content. Remediation is accomplished using the Python programming language’s lxml package and validated with a custom schematron file. This lightning talk will present some of the changes required for the remediation and review how lxml and schematron automated the process.

NOTES: UVA works in MODS with many input standards. DPLA uses their MAP for display in their portal. Worked with DPLA staff to implement changes. Used LXML Python module: has been used for several changes, including making metadata source codes consistent. Also used Schematron to help validate XML patterns that they require for their implementation. Still working with DPLA to get their data in, almost there.

Presentation Title: Overcoming the Challenges of Implementing Standardized Metadata Practices in a Digital Repository
Presenter: Sai Deng – Metadata Librarian
Affiliation: University of Central Florida
Abstract: While implementing standards in cataloging digital collections is often a Metadata Librarian’s conscience or inner desire, sometimes it’s a challenge to do so if a system is not built to accommodate such standardized practices. This kind of dilemma is not uncommon in the metadata and digital repository arena. This presentation will address the various challenges in working with metadata in digital repositories such as, name authority control for authors, departments and colleges, type values selection, keywords and subject choices, whether to add linked data URLs to various fields in the records and data discrepancies in harvesting data into the OCLC’s Digital Collection Gateway. Sometimes trying to follow controlled vocabularies or standardized metadata practices seems to be at odds with what the system can accommodate or what many non-catalogers prefer. This presentation will discuss how the Metadata Librarian, Digital Initiatives people and other librarians work together to make careful, practical and conscientious choices.

NOTES: Looking at workflows, involving different kinds of staff to do authority work. Issues of differences between what repositories want, lack of standardization. Closed systems also a problem, with needing to have vendors make changes to fit a repository’s wants. Possible solutions: work with others to establish templates with local standards baked in.

Presentation Title: Using MarcEdit to retool existing MARC records of paper maps for use in an online geoportal
Presenter: Tim Kiser – Special Materials Catalog Librarian
Presenter: Nicole Smeltekop – Special Materials Catalog Librarian
Affiliation: Michigan State University
Abstract: The Michigan State University Libraries recently joined the Big Ten Academic Alliance Geoportal, a consortial online discovery tool for maps and geographic data. Contributing our scanned paper maps to the geoportal required submission of metadata suitable for the generation of ISO 19115-compliant records. To accomplish this, we devised a workflow using MarcEdit to convert our existing MARC records for paper maps to MARC records for digital maps — which could then be delivered to the geoportal as MARCXML records. This lightning talk will outline our considerations for the project and the steps taken to accomplish it.
NOTES: They convert MARC into the ISO 19115. The MarcEdit workflow changes a number of fields in MarcEdit, while some changes have to be done manually (e.g. 776, 6xx, other fields). They have 44 maps into the consortium. Lessons: adhere to provider-neutral guidelines; remove FAST headings and let OCLC generate new ones; add 347 next time.

Presentation Title: Metadata Migration to Leverage Linked Data in an Institutional Repository
Presenter: Brian Luna Lucero – Digital Repository Coordinator
Affiliation: Columbia University
Abstract: This talk will present the project of migrating records to a new cataloging tool for Academic Commons, Columbia’s institutional repository, with an emphasis on metadata modeling for the new application and transformation of the subjects for all records from the ProQuest vocabulary to FAST. Over the last year, Columbia University Libraries has supported development of a new cataloging tool, codenamed Hyacinth, for digital collections in order to unify the workflows of several departments and ease the demands for maintenance of multiple platforms. Hyacinth also provides an upgrade over older tools by operating on Hydra architecture and incorporating linked data at its core. Creating one tool that suits the cataloging needs of different departments and projects presented its own technical challenges, however. Hyacinth serializes records in MODS XML, but was designed to be scheme-agnostic. Achieving this aim required input from metadata experts familiar with the various projects and materials that would be handled by Hyacinth. Normalizing labels for names, genres, academic units, and subjects across numerous projects and departments also presented a challenge. This led to the creation of a URI service that is integral to Hyacinth. The URI service can pull information from external authorities as well as mint local URIs for entities not identified elsewhere. The migration of Academic Commons records also required a transformation of subjects for approximately 20,000 records to the FAST vocabulary in order to capitalize on Hyacinth’s linked data architecture. We used OpenRefine and a mapping table to replace ProQuest subjects with equivalent FAST terms and add FAST URIs to the records. We also piloted text matching processes to see if any can automatically suggest FAST subjects that match keywords in abstracts. These experiments have produced mixed results.

NOTES: They worked on a second custom tool for cataloging. The current version allows for batch remediation that goes from CSV to JSON. One use: adding big batches legacy dissertations, another: batch changes of department names in repository. Challenge was reconciling used vocabularies against FAST, going from ProQuest dissertation headings to FAST. Have used OpenRefine identity recognition, useful for geographic, FAST matching, names. Ahead: other things not covered before. Challenges: very bespoke local tool that has to meet all needs.

Presentation Title: Metadata Librarian’s Little Helper: OpenRefine Reconciliation Services
Presenter: Greer Martin – Discovery & Metadata Librarian
Affiliation: Illinois Institute of Technology
Abstract: OpenRefine has many vocabulary reconciliation options, not only with Library of Congress Authorities and VIAF, but also with homegrown data such as a local authority file. With unruly legacy metadata, reconciliation was a major chapter in the story of our records migration to ArchivesSpace. Taking a systematic approach to our vocabulary reconciliation and using OpenRefine’s reconciliation services allowed non-catalogers to assist in this crucial stage of metadata cleanup. This lightning talk will explain how two OpenRefine reconciliation services
were incorporated into our migration workflow, with special attention paid to Reconcile-csv, which resolves to a CSV file.

NOTE: Moved systems but previous headings lacked authority control. Used OpenRefine reconciliation tool. Started with batches of 100 records. "Pretty good" match rate, about 50%, either direct matches or suggestions of matching. Under a minute for 100 records, about an hour for human cleaning up of suggested matches. Ended with two document type: one with LC-reconciled names, and the other with local names; created master CSV of all names. On master CSV, matched unmatched names; said was easier than against LC. Reconcile-CSV can work with OpenRefine for this final reconciliation step.

Presentation Title: Git a Grip: Using GitHub to Manage your Metadata Application Profile
Presenter: Anne Washington – Metadata Librarian
Affiliation: University of Houston
Abstract: Local Metadata Application Profiles and input guidelines are always evolving. GitHub provides a simple way to manage metadata documentation with the added benefit of versioning. This allows metadata specialists to see changes in practice over time. Learn how University of Houston Libraries is using GitHub to create and manage their Metadata Application Profile.
NOTES: MAPs change quickly and there are needs to make changes. Also need to track changes, so GitHub provides a good solution for the format of their data dictionary. Theirs is an HTML page the versions of which GH lets you manage. Uses desktop GitHub and then commits to online GH. Can comment in a note with each change.

Mike Bolam: Still looking for presenters for a summer preconference presentation on diversity/equity/representation in metadata. Also looking for program content on metadata migration workflows. They have 2 good ones for each, but would like a couple more.
Discussion afterwards: some issues with vocabularies for defining type of resource (e.g., PowerPoint). Is it type, genre, format? MODS is insufficient to describe datasets.
One person reported that their repository had 100+ datasets. They were hoping that patterns would emerge as they add more content, but they're finding that there's a huge amount of uniqueness to each dataset.

Business meeting:
Includes MLA report (At end of the report for this session)
CC:DA report. Evolving towards a new structure for representatives, with only one for North America. Working to sync RDA toolkit with the Open Metadata Registry. Working to replace language within the Toolkit. Three R project to be complete by April next year, mainly to accommodate IFLA-LRM. "Four-fold path" working from levels of description from free-text to URIs. Monday. PCC has task force on coding gender in authority records; webinar to come.
Some work correcting information on the blog.
They have 3 openings in the MIG:
Vice-chair/chair-elect
Program co-chair
Secretary for 2017-2019
ALA has put out a "conference remodel" to improve user experience and reduce cost. Changes include unifying program submissions to a single form; reduced number of program slots; each ALA division will have a set number of slots for trending topics. New Orleans 2018 will be the first to implement. Must begin submissions 10.5 months before annual. Really impacts agility. Moving to move programs to convention centers; other meetings to hotels. That last part about hotel meetings may change. One shift is that meeting times would shrink to one hour.

Discussion about Metadata Blog
Cleaned up content. Now thinking about developing content not up at ALA Connecct. Maybe add content related to program slides? Use the blog to market upcoming presentations? Profiles of presenters, with descriptions of what they really do at their jobs? (Metadata Librarian positions are all over the map.) Easy Google Form that feeds to Blog Coordinator?

Michael Bolam - ALCTS Interest Group

Report from Music Library Association for Metadata Interest Group
Prepared by Jim Soe Nyun, Chair Encoding Standards Subcommittee, MLA liaison
Past reports have been pretty heavy with information on MARC development, but there’s been only one small proposal that we currently have in the works, a fast-track proposal to make one of the MARC fields repeatable, Field 384, Music Key.
Last year I mentioned that MLA would be pulled in to the Performed Music Ontology project, one of the several components of the LD4P Mellon grant. This particular project is a on quick timetable, and its report should be out in just a few months. MLA has two people including myself who have been directly involved with the PMO project. And MLA has formed the Linked Data Working Group, LDWG, known affectionately as “Ludwig,” which has formulated a number of use cases that have gone on to the PMO group. LDWG also has been involved in providing feedback on some of the early work coming out of the LD4P project, and the group has also participated in helping critique ontologies that have significant features that might help us look at how to model events. If you’d like to hear more on the project, Nancy Lorimer will be presenting an overview and update at the LC BIBFRAME Update Forum coming up at 10:30 this morning.

LC BIBFRAME Update
Sunday, January 22, 2017
10:30-11:30

Update on Recent Developments--Sally McCallum
Detailed specs for MARC to BF almost completed. Will be published within a month, subjects to frequent updates. Conversion program also will be made public. MARC-to-BF display tools ind development, with MARC XML on one panel and BF on another. Metaproxy has BF markup among other options. Casalini and ExLibris and SIRSI/DYNIX are experimenting with BF. MODS to BF conversion specifications done.

LC Plans for Production Pilot 2--Beacher Wiggins
Pilot plans have been delayed. New kickoff date will be May/June. 45 staff from the first pilot
have continued cataloging in BF since the pilot ended. Ongoing meetings with these catalogers. One big gap with P1 was that there was no way to edit/correct BF cataloging. This new pilot will have converted the entire catalog converted. BF editor also to be updated. All formats, languages, scripts. Will resume with original participants, and will expand to 80 or so staff participating.

Music Development for BIBFRAME in LD4P--Nancy Lorimer

[An abbreviated version of a longer talk to be presented later Sunday at the PCC Participants meeting]

Gave overview of the Linked Data for Libraries grant project, and how Linked Data for Production follows as the next phase. Several sub-projects constitute the larger grant. The Performed Music Ontology project is one of two Stanford projects.

Some work on developing changes or additions to BIBFRAME vocabulary 2.0. Includes defining many types of titles, which were designed as subclasses of bf:Title, and not bf:VariantTitle.

Working on building into model elements from existing vocabularies, including those in the RDA Registry (RDA terms, unconstrained properties), MARC Relators list.

Showed work on developing a model for thematic index numbers, with elements to include: the number string itself, the components of the number (prefix, number, NumberPart), the agent responsible for assigning the number, and the work where the assigned number appears.

Work on incorporating outside vocabularies involves emphasis on developing “individuals” in the data model; they cannot be subclassed and can represent intersections of classes.

Bringing MARC forward to BF--Wayne Schneider, Index Data

Worked with LC to develop legacy MARC data into BF 2. 19M MARC records is a lot but not really "big data," so big data tools might not do much good. 2B triples may result from converting LC’s database. Tries to use things like VIAF to link MARC content to authorities.

XSLT1.0 converter tool for first static conversion step. Conversion is done through a series of lookups (ca 900).

Future work, configuration, develop schema for MARC to BF.

Future of libraries is Open (FOLIO). Work on interfaces between modules that can be developed by the community. "A community collaboration to develop an open source Library Services Platform (LSP) designed for Innovation." “Anti-ILS” in structure.

OCLC’s work on works--Roy Tennant, presenting for Jean Godby

Talked on WorldCat Work: You can’t rely on them, they will be regenerated using a new algorithm. Begins with clustering by author and title. Subdivides next by genre and resource type. Extract content-oriented fields.

Discussion on works:
VIAF works: OCLC using VIAF works to extract multi-level description. Issues with modeling work + expression.
OCLC and LC Works comparison, "super work" a concept that likely needs to exist.
PCC Works Task Force: Looking at definitions of work.
URI Task Force: Worked on how to insert work URIs into MARC records.

Questions
Did Index Data produce a schema for record conversion as a byproduct of their work? No time to do. LC is finalizing a maintenance contract with Index Data.
Is WorldCat clustering going to be used for the work being done on works? Don’t know. OCLC Research realizes there are problems.
Will messy things like ISBD punctuation in MARC be removed in the conversion? Only if they’re part of the label and important to retain.
(Mark Scharff): Will there be a way to describe that a work isn’t in a thematic catalog? Not in the current model. Maybe something to fit in.
On the early thinking on things like Expression in their model, a new work, a subclass, something else? Most work is on relationships, not really on how are these things called. Please add language information so that inferences can be made to help out with drawing conclusions on what other records might not have coded.

Metadata Standards Committee
January 22, 2017
1-2:30 pm
Georgia World Congress Center, A303

AGENDA & NOTES

Welcome and introductions

Visit from ALCTS president Vicki Sipe

NOTES: President and President-Elect came in to discuss once-every-5-year review. The committee has presented a report with ways its charge could be redefined as the metadata world has shifted. The process would be that after the renewal goes through, then changes could discuss. Discussions about process about how the group can publish content, and a mention that LibGuides is now available to use. Much concern that the new ALA conference structure, with its reduced meeting slots and the need to plan meeting content much farther in advance, would harm the effectiveness of groups such as this committee. Some discussions that ALA/ALCTS would like the committee to use more of the official communication channels, and comments that current communications structures with an independent blog came up in response to difficulty with making content available through official channels. Encouragement to again try using more official channels where possible, that things had changed since the Metadata Policy Committee first tried to use the official tools.

Question from the community - Discussion of metadata needs related to accessibility of resources and bibliographic metadatathttp://bit.ly/msc_accessible

NOTES: Eric Mitchell was interested in developing a "HathiTrust of accessible content." Questions about what is out there in the way of encoding standards. I mentioned that MAC is taking up
DP2017-03, devoted to accessibility information. Questions about how vendors might participate in this. Units that have to do accessibility remediation might have good input to this.

Ideas for new projects for the committee this year

NOTES: Ideas include how to get the principles for metadata standards better publicized; maybe track where the principles get cited. Any thoughts about how to decide when MARC to LD conversions are "good enough." Time to re-ask the question, "What is metadata quality?" How to improve publisher metadata quality?

Commenting on draft standards

Upcoming calls?

NOTES: Three people will be working on the International Council on Archives' Model on Archival Model. (The FRBR for archives.)

Programming at MIG at Annual on putting our Principles for Evaluating of Metadata Standards into practice

NOTES: Idea to present at MIG how the Principles have been used. Also, we should start thinking about programs for Annual 2018.

Work on new draft charge

NOTES: How to development a new charge. Start with current charge and modify. Worked through some options and will develop a draft afterwards.