Discussion Paper No. 2018-DP05: Adding Institution Level Information to Subject Headings in the MARC 21 Bibliographic Format

ACTION: Attendees imagined several scenarios where metadata provenance could be both useful and problematic. Some envisioned a scenario with multiple competing terms, and institutions could possibly end up appending $5’s for everyone who might be using a term. In the end these concerns seemed like they could be dealt with via best practices. This will probably come back as a more formal proposal or paper.
Discussion Paper No. 2018-DP06: Versions of Resources in the MARC 21 Bibliographic Format


ACTION: Good support for this paper’s aims, with consensus that there are situations when recording a version of a resource is important. There was discussion of the various options offered, with a final determination that the recording method chosen should be designed to accommodate controlled vocabularies for states and versions of resources.

OCLC Linked Data Roundtable
Saturday, February 10
10:30 AM - 11:30 AM

Roy Tennant leading panel on OCLC linked data projects
#oclcldr

Sarah Newell, OCLC
Linked Data Prototypes

- Last fall collaborated with UCD, Cornell and a Montana State Library about LD use cases, building on previous LD prototype projects. Develop a Reconciler, Mint, Relator to connect legacy data to LD entities, mint new entities, and one further function.
- Working toward a centralized hub, part of an Entity Ecosystem. Developed database with a reconciler on top; next step is something to allow editing of entities. Used Wikibase for display/search, an out of the box tool for the project. Lesson: good communication to develop project was needed; needs more than just converting records, need to edit; scale matters, that enough data is needed.
- Next: creating and editing entities; going from feedback to enhancing functionality; expanding project out from initial partners. Looking for partners: contact newells@oclc.org.

Sally McCallum
Talking about MARC, converting 19M MARC records into BIBFRAME

- Looking at work records: In MARC authority of bib description? Currently in bib to permit “uniform title” transformation. They have a continuous flow of incoming MARC.
- Wondering about what to add to authority format that is currently in the bib format.
- Needs: 100 author/title strings that match standard strings
- Place to records notes about formulation of the string, e.g. 670
- Place to record title variations (4xx)
- Place to record to all versions of the title, e.g. common title i law.
- Adding 400s, 381s for e.g. translator and 377s for language, also 500s for related agent, e.g. translator.
- Looked at how they are used, by catalogers, by end users.
- What’s missing from the bib format: statement that a bib is for a work, series control fields, variant title fields, related title fields, related agent names.
- The above MARC tags don’t relate to direct transcription from one format to another, authority 5xx may end up in 7xx.
- Bib content missing from authorities: things like 520s, topics, etc, how much is really needed.
- LC doesn’t make work records for everything. For all? For when there are 240s? A system should do this work, extracting from existing metadata.

M) Hahn (Myung-Ja), Illinois
Linked data for user services

- Why are we doing this?
  - As a producer, to make local resources discoverable on the web.
  - As a consumer, provide web content to local users.

- Emblamatica Online project 2015
Portal for Renaissance texts and images, digitized resources from ca. 1400 books and 28K emblems. Uses VIAF to provide additional information, e.g., gender. Generates a “More info” button from VIAF in the Emblamatica.

Using IconClass to user services, has 4 language options (supporting international collaboration on project), terraced vocabularies at different levels.

Other project: LOD for Digitized Special Collections

18 mo exploratory study funded by Mellon Foundation

Special collections are silos, aimed to map schemas to LOD to expose resources.

Theatre collections generating knowledge cards for entities, enabling browse by related resources. With Wiki links. In item page, developed a sidebar of related resources. Can generate link clouds, that users can annotate.

Lessons learned: Contextual information can help users with cards that leverage LOD sources; this really is resource-intensive. Other things to do with LOD are still to be explored.

Questions: How is an entity ecosystem not a triple store? Ecosystem is a format-agnostic database, more functionality than a triple store.

In LC’s conversion, are they dropping data in the conversion? The conversion specs include some not-applicable data and other things that they chose not to encompass.

“Work?” Lots of definitions, but Sally was talking about a BF work.

Is there proof that all this LD work is helping users? The project includes assessment tools, and they will be continuing work next month. They are using Google Analytics. Results of analysis will be published soon. // Sarah Newell mentioned they have nothing to add // Sally acknowledges things are important.

Are there stats on LD projects at Illinois? No direct numbers ... yet.

In addition to improving the user experience, are there things being done for machine discoverability? Using Schema.org to exchange SEO. Both Illinois projects used such schema.org.

Sally: How much of the LC work is manual versus automated. The projects presented are actually talking points, not work conducted.

Cards, etc., point outwards. Potential of losing lots of users. Confidence that users found the pages and that they’d be back.

How bib is the ecosystem? > 1M items. Small subset, could easily exceed 1B.

Are LC’s MARC discussions likely to head to mARC changes? These are just investigations. They hope to repurpose what’s extant. Exploratory. Changes to make LOD easier.

What to do with subject string conversion when there aren’t exact subject matches. LC converts them now. They use labels when there aren’t matches.

American/British items published w/o UTs, how to handle? Just weighing pros and cons of working the bib file.

If their project is successful what will you gained from it? MJ will publish scripts and everything else used in their project. (Github page.) // LC [aside: 15 years of a hybrid environment] may have some discussions about how to enhance description. // Sarah, a better understanding of what problems they’re solving.

MARC Formats Transition Interest Group
Saturday, February 10
3:00 PM - 4:00 PM
Program on ALA Connect: http://connect.ala.org/node/248329

Deborah Shapiro, moderator

Theme: Local Tools, what’s being done locally with linked data

Barbara Bushman, NLM
Preparing for the Future: National Library of Medicine’s Project to Add MESH RDF URIs to its bibliographic and authority records

Adding URIs doesn’t hurt current MARC implementation and prepares the catalog for the future. Focus on names and corporate bodies, but started with MESH, a product they own and understand. Each string must be controlled, unlike the open structure of LCSH, which allows for far more (almost unlimited) string combinations. They have 850K MeSH records and add 350 bib records with MeSH. Would like to add $0, and has dispensation to add them to BIBCO records, even though it[s not yet PCC practice. Started with MarcEdit MARC Next tool, had 100% matching. Voyager didn’t let the
records be update without taking them out and reinserting. Made their own program to add RDF URIs for MeSH, first to authorities then bibs. The process took 36 hours for their catalog. They no do this for all current cataloging. Worked on community impacts, notifying users of their records, including the giant update that included all their database. CIP display is ugly, so they suppress the $0. OCLC didn’t want them since they figured out adding MeSH, though not in actionable form. Realized they have a very clean database, and then further cleaned up 4100 MeSH errors in their catalog.

- Next steps: Collect feedback. So far only one library has inquired what these $0s are good for. Work with MarcEdit, Authority Toolkit, maybe other tools. Work with Casalini to publish some NLM data as linked data.

Sally McCallum, NDMSO, LC
Transforming a Catalog—Converting MARC records to BIBFRAME
Jodi Williamson could not come to give this report, Sally presented her slides.

- Transform components: BIBFRAME model revised to 2.0, revised all specs, including conversion programs, also infrastructure improvements (more servers, MarkLogic. Added Item information and Events added to model. New specs include retaining extinct LCCNs, ability for Duration to repeat. Specs also got URIs AND labels from code lists. (These also can feed into the editor and keying tools.) 18M MARC records converted to BF Works, Instances and Items; 1.2M uniform title works converted.

- Developed Comparison Tool

- LCCN entered into it will show how a MARC record is transformed into RDF.

- Next step, Match and Merge

- MATCH: Discover all the Instances (manifestations) for a Work; Discover if Work descriptions already in file (from MARC title authority records), if not creat Work. MERGE: Merge subjects and other info into record.

- Title AF => Work record (BF work)

- Instance subjects merged into single Work record.

- Instance records link up to Work and down to Item.

- LC BIBFRAME Editor: "Not flawless, does a pretty decent job." Catalogers appreciate the type-ahead features. Some drop-downs. RDA rules hotlinked to element labels.

- Entire workflow: Has a process to double-catalog things, MARC and BF. 200+ non-pilot catalogers, 60+ Pilot catalogers; 19.2 Milligan Work, 23.7 M BF Instances. RDF file of >4 billion tripes.

- Close to offer LC’s BF file for others to explore; mapping from BF to MARC to simplify their workflow. Do works records in the bib file. Looking at more records with transliteration.

Brian Rennick, BYU
Enriching Author Records with Linked Data

- Tyler Ashcroft wrote the code for this project

- Talked about enrichments that live OUTSIDE the catalog; they have a federated search, Symphony from SirsiDynix

- Inputs include MARC XML, XML from Wikidata (using SPARQL)

- Processed with Apache Jena in Java, cached in Mongo database; Author landing pages linked from federated search. These components come from author landing page.

- They tried out Zepheira’s Identifier Services, but found LC data to be easier to work with

- Wikidata Enrichments: Birth/death place/date with geo coordinates, birth names, other facets; also pulls in IDS from other sources including IMDB.

- Followed Wikidata and Blazegraph requirements to process Wikidata files

- What’s next? This has been an experiment. Want to publicly share landing page for authors. Next to subjects and genres?

Questions (several asked, selections below):

Does NLM add component links to compound strings? No. Not in MARC bibs, but is in RDF, which includes allowable qualifiers.
Faceted Subject Analysis Interest Group
Saturday, February 10
4:30 PM - 5:30 PM
Notes on ALA Connect: http://connect.ala.org/node/273985

Kelly McGrath, University of Oregon

- LCSH difficult to use unless keywording for browsing, and then would need a system that does stemming (saxophone/saxophones)
- MARC 048 was never fully developed
- LCMPT better implementation for medium of performance, and its uptake could be enchanted by the new Strawn/MLA MOP Toolkit.
- Developed at Ball State, with Sue Weiland, a search for chamber music that used the 048. They generated 048s using the richer IAML instrument vocabulary. Also added local $z for Form, e.g. sn (Sonata) or chbr (Chamber music).
- Oregon has Alma to Primo workflow, and Primo has ways to transform MARC in interesting ways. They've developed something similar to the Ball State project that uses MOP statements
- They've developed a medium of performance statement facet, though it’s not perfect.
- They've also developed something for numbering in the 383. Leads to messy displays when there are multiple $a's or $b's.
- They also mine 700 strings for title elements.
- Also a facet for subject strings, they find easier than LCDGT, partly because of prevalence in data.
- Retrospective conversion might be needed to make these facets more useful because bib records are not consistently coded with facetable data.
- Another problem is that there aren't ways to navigate within hierarchies.
- Questions about whether to record this information in authority or bib records. More contributors to bib records than authorities; authorities have the promise of less work/maintenance.
- Messes can happen with compilations, collaborative works.

Diane Vizine-Goetz, OCLC
Update on FAST

They’ve developed a tool for proposing FAST headings. For names, corporates and topical. Short turnaround. Personal and corporate names aren’t added automatically, only when they appear as subjects because most names aren’t ever used as subjects.

Users must use existing LCSH strings. Can assemble components into a longer string. Doesn’t prevent bad string construction. But the plus that headings can be done at the point of cataloging is a big plus.

New to FAST: event headings from 111 fields into 147 fields, using new field. Structure of heading is sub fielded and different from 111 and LCSH forms.

There was a FAST survey front he end of last year. Results have been tabulated and will be distributed soon. Preview of results: Need tools to support ongoing maintenance; production tool for FAST heading look-up; seamless addition of FAST to records at the point of cataloging.

FAST does include a de–duping process, e.g. Biography and Biographies. But they haven’t done it in a while wile.

Does FAST record what LCSH record it was based on? Sometimes, when there are straightforward correlations, but not when FAST came about via rules.

Question about whether geographic will be harvestable from NAF. Not just yet. It’s more complicated to derive FAST geographics. They are still strongly linked to what’s in LCSH.

Question that non–English materials are not well served by LCSH. A future development opportunity.

Metadata Interest Group
Sunday, February 11
Title: National Strategy for Shareable Local Name Authorities

Abstract: Libraries create local authorities to serve a variety of purposes, usually within an institutional context. It is becoming increasingly evident, however, that identities have much greater potential value if they can be shared. The IMLS Shareable Authorities forum brought together representatives from a wide range of stakeholders to explore themes including minimum viable specifications, data provider obligations, and reconciliation as a service. The objective of the forum is to identify services and practices that will be needed, and assumptions that will have to be made – or changed – to allow authorities to work across domains and at scale.

NOTES:

- IMLS Shareable Authorities Forum
- 2 meetings held at Cornell and LC in 2016-17. IMLS grant to Cornell University, partnered with LC, OCLC, PCC, ORCID, CNI, SNAC, BIBFLOW, Stanford, Harvard
- Deliverables: Report and reference model in progress, a few more months before final report
- Why local authorities? Interest in linked data from supply end, ability to scale
- Why a forum? IMLS has a grant program for forums, less stringent than other grant models; Cornell was already working on these issues. Opportunity to share knowledge, even if a solution to problems won’t be a product of the grant
- “Linked data changes the game” Even with increased exposure and sharing, an authority file needed cross-platform
- Related projects (Inter Alia)
  - IMLS Western Name Authority File Project
  - PCC
  - OCLC organizational Identities in ISNI report
  - Others...
- 40 people involved in forum
- What they learned: Discussions focus a lot on use cases or institutional mandates. Data models can evolve from competing organizational needs. Persistence an issue; heavily in the experimental realm
- Workflows
- Technical needs
- Social/organizational goals
- Attendees at forum come from different places:
  - BL interested in developing workflows for expanded collections
  - OCLC, Getty interested in requirements for aggregation and LOD publishing
  - Publishers want to use identities, don’t want to manage
  - Self-registration (ORCID)
- Modeling issues:
  - Provenance (important in archives community)
  - Preferred labels (big in library world, but can impede sharing LOD; ISNI ignores labels)
  - Granularity
Social, organizational issues:
  o Licensing, business modeling (solutions for now aren’t always maintained)
  o Privacy, confidentiality
  o Sustainability
  o Governance

Outcomes, directions

Algorithms for matching are important. Maybe everyone should share these. Will this really happen?

Develop a “Minimal viable product”

Reconciliation service

How to provide data in a way that’s suitable for aggregation

Emphasis moving to corroborating data

Best practices may be domain specific

Reconciliation as a service

A stack of software and data would:
  o Harvest authorities
  o Work for many use cases
  o Deal with wide range of quality
  o Work with degrees of confidence

Responsibilities of Data Providers:
  o Provide provenance
  o A voice redundancy
  o Code disambiguation data in machine actionable forms
  o Allow iteration based on reports

Responsibilities of Aggregators:
  o Err on the side of duplication than mis-conflate
  o Provide unique persistent cluster identifiers
  o Record provenance of individual data elements
  o Use but respect confidential proprietary data

Ayla Stein
Metadata Librarian
University of Illinois at Urbana-Champaign

Title: Developing A Framework for Measuring Reuse of Digital Objects: Project Update at the Metadata Interest Group, ALA Midwinter 2018

Abstract: Content reuse, or how often and in what ways digital library materials are utilized and repurposed, is a key indicator of the impact and value of a digital collection. Traditional library analytics focus almost entirely on simple access statistics, which do not show how users utilize or transform unique materials digital collections. This lack of distinction, combined with a lack of standardized assessment approaches, makes it difficult to develop user-responsive collections or highlight the value of these materials. The grant project, Developing a Framework for Measuring Reuse of Digital Objects, an IMLS-funded project (LG-73-17-0002-17) by the Digital Library Federation Assessment Interest Group, is working to address this critical area. This presentation will illustrate the variety of ways digital library objects (including metadata) are being reused; share the results of the grant team’s work, including preliminary findings from the initial survey results as well as in-person and virtual focus groups sessions. The presentation will conclude with the
team's early findings and will engage the audience to contribute their feedback on the project and deliverables.

NOTES:

- Full team includes 6 organizations
- Started out with white paper: Surveying the Landscape: Use and Usability Assessment of Digital Libraries (2015); 2017 National Leadership Grant from IMLS
- Goals: ID sustainable and vetted assessment techniques that can support many kinds of digital collections / Support cultural heritage organizations ... / 3rd goal
- Data collection:
  - Initial survey
  - Focus groups: in-person, also virtual
  - Final survey
- Deliverable:
  - Web site so far
- Use vs. Reuse: Difference
- No community definition
- Developed working definitions
- Use: Discovering and browsing objects in a digital library, often described as clicks or downloads, without knowing the specific context for this use.
- Reuse: how often and in what ways digital library materials are utilized and repurposed. In this definition, we do know the context of the use.
- How does metadata fit into this?
- Metadata = Data
- Collecting all types of digital data
- So far:
  - Initial survey complete
  - In-person focus groups, 1 round done
  - Virtual focus groups, 1 round down
- Final survey, informed by focus group data, not yet designed
- First survey, concentration on how do cultural heritage organizations assess digital library reuse? What resources would help with reuse?
- Some survey numbers, 409 responses to 19 questions
- 80% collect use statistics
- 40% only collect reuse statistics, citing lacks of methods, lack of staff
- Highlighted ideas
  - Assessment metrics must demonstrate impact
  - Need training on how to gather stats, how to advocate for resources via metrics
  - How to deal with patron privacy
  - People are interested in metrics, would like best practices for what methods to use
• Takeaways & next steps
  o Use survey results as benchmark for community progress
  o NO FREE TEXT in survey design (audience suggestion: test your survey before conducting it)
  o Collect and analyze focus group data

  https://reuse.diglib.org

Slides on ALA Connect, linked to Metadata IG Blog

BUSINESS MEETING

• Welcome

• Officer Reports:

• Program for ALA Annual, keynote (Philipp Schreur) on topic of LOD, have a call out for additional speakers

• Blog for the IG: ALCTS does backups for the blog; 144 posts back to 2007; 30 legit comments out of much spam

• Elections at ALA Annual, encouragement to run for office

• Review of minutes:

• Liaison reports

• MLA (brief) report:
  o 2 members from MLA were part of PMO project (component of the larger LD4P Mellon Grants) which issued several draft papers this year about features of the Performed Music Ontology ("An extension of the BIBFRAME ontology for describing performed music, both for mainstream and archival performed music collections") including a detailed, complex one on medium of performance. Available on BioPortal (https://bioportal.bioontology.org/ontologies/PMO) and Github (https://github.com/LD4P/PerformedMusicOntology).
  o MLA’s Linked Data Working Group (LDWG) about to start up again after brief hiatus, main task to examine Performed Music Ontology, including an effort to compare its products against use cases developed by LDWG.
  o MARC changes:
    o Joint MLA–OLAC task force to look at what to do with the 33x and 34x MARC fields.
      o 33x fields will be left in their current compact form, following what the Library of Congress stated would be its policy.
      o 34x fields will be exploded out, one subfield per line, with source vocabulary for that subfield cited; opportunity for the identifier for the term to be included in the field.
  o Discussion Paper on adding $3 to 5 MARC fields
  o Fast-track changes to Field 384, Key, made repeatable, added $3, Materials Specified

• CC:DA report: Kathy Glennan incoming Chair; RSC NARDAC documentation; RDA Toolkit redesign previews issued; June release will be thin, 1 year duration for keeping the old Toolkit alive; another meeting coming up tomorrow, program on new directions for CC:DA

• Metadata Standards Committee: Mike Bolam, evaluated NISO standards about evaluating standards; looked at DPLA metadata application profile; they realized that there are high barriers for the committee to try to make good evaluations of metadata standards; looking to redefine direction, using it as a forum for how vendors and entities are dealing with metadata, including issues of diversity and inclusiveness. Meeting Monday 1-2:30. At annual will have a program: Metadata Experts Panel, addressing current challenges in metadata issues.

• Discussion:
  o Job openings at NYU, University of Utah, 2 positions at Smithsonian Institution

Library of Congress BIBFRAME Update Forum
Sunday, February 11
Library of Congress Pilot 2, Sally McCallum NDMSO, LC

- Beecher Wiggins not present, only 15 of the original 200 staffers from LC got to come
- Pilot at LC is in full swing, after starting out in September
- September saw a European BIBFRAME conference; they’re still exploring, just like in the US
- BF Pilot, June 2017 start, 1 year period for evaluation
- 60 catalogers involved, books, serials, maps, music, moving image, rare, sound
- Worked on training for Pilot participants in the new input tools, included some RDF training to doublecheck RDF that was generated from their cataloging.
- Base file: converted 18 million MARC bibliographic records to BF Works, Instances and Items; 1.2 M uniform title authority records converted to BF Works, merged records into 19.2M Works, 23.7 M Instances
- Continuous review and frequent reload of base file based on errors in ingest when detected
- New input tool
- Editor has efficiency features like dropdowns and lists, hot links to RDA rules; ways to view resulting RDF
- Challenges with synchronizing Editor RDF with RDF of converted records; profile elements and MARC derived elements; complexity of RDF.
- Want to explore: improving Editor, validating BIBFRAME from MARC; download of BF file; import BF RDF from an external source; experiment with extension profiles; conversion for BF to MARC; ?Works in MARC Bibliographic file rather than Authority file; reduced transliteration in descriptions
- Community Explorations: www.loc.gov/bibframe has components: BF vocabulary, MARC to BF conversion.

Folio and BIBFRAME, Sebastian Hammer, President, Indexdata

- BIBFRAME IN FOLIO: Thinking about bibliographic metadata in a new LSP
- FOLIO a collaboration that began within EBSCO, worked towards developing a community library platform.
- FOLIO manifesto: 1) a true platform is like an operating system that can evolve and grow over time 2) an open platform is common property, all have rights to use and improve 3) a true and open platform has the potential to gather the library community broader than any closed system
- FOLIO consists of a UI toolkit (Stripes) and API gateway (Okapi)
- FOLIO can include other components, e.g. different UIs or different API gateways
- All being tested now. To work it needs to deal with a broad spectrum of metadata input. No native BF triple storage yet. Apps looking at things like metadata and resource management and circulation
- Modeling includes a Codex that can access metadata in multiple ways
- Came up with a data model that uses much of the BF model, uses work, instance, item/holding levels in the Codex, an abstraction, this would allow use of MARC, BF, or just the 40 elements in the base FOLIO model—or some other TBD standard
- Not much for libraries yet, but maybe 812 months from an ILS

Alma, Linked data, and BIBFRAME, Amy Pemble, Product Manager, ExLibris

- Linked Data Implementation at Ex Libris, Amy Pemble
- Has worked with Linked Open Data Working Group (2011)
- 2017 worked with Harvard in a BF implementation. Now all Alma institutions can publish their collections in BF as of last December.
- Can publish their collections in RDA/RDF as well
- Linked data APIs provide API endpoints for JSONLD, RDA/RDF BF.
- Includes linked data views of RDF, and can display MARC and BF side by side.
- Primo can take URIs and provide discovery opportunity from external resources
- Alma will continue with MARC, but will search linked data endpoints from within the editor, as well as catalog in BF. Can export data with schema.org elements.
- Summon will be linked data aware later this year
- Alma can export linkable data, still looking at options for reconciliation, a process that still requires much manual work

Achievements of 2016/2018 LD4P Project, Michelle Futornick, Program Manager, LD4P

- Recap of LD4P grant
- Focus on metadata creation model of LD lifecycle
- Recap of the component projects looking at non-text formats
- Recap of data modeling efforts, partnering with domain expert communities, including Stanford developing the Performed Music Ontology
- Example: in ArtFrame/RareMat groups, use case, Find all the resources connected to a specific award across various domains.
- Links to models at LD4P.org
- Tools development: Developing "editors" that actually are for content creation; comparing BF Editor, VitroLib, CEDAR to come up with desirable features for an editor
- Looked at Stanford Trace Bullet project, looking at vendorsupplied copy, showed an implementation that includes MARC and BF metadata; developed "reactive pipeline" which allows for constant updating of content as source material evolves
- Community meetings in April 2017, will have another in May 2018.
- "We are in our infancy as far as linked data tools."
- Stay tuned.
- Developed Biblioportal (bibliographic.ontoportal.org) a catalog of ontologies, an outgrowth of Bioportal, developed for libraries.
- LD4P2 "Pathway to implementation" the next outgrowth of this project
- Developing LD editor sandbox in collaboration with PCC. Expand on workflows to go from MARC to BF. Want to build out to develop a self-sustaining model to maintain tools. Identifier management. Enhancements to Blacklight to show off LD components.

BIBFRAME and OCLC, John Chapman, OCLC

- OCLC's Work on Works
- Why: works are critical for clustering; there are common use cases; interest arising from ongoing cooperation with OCLC and LC and there's community interest.
- PCC SCS/LDAC Task Group on the Work Entity, published in Fall
- Cornell leveraging OCLC Work IDs in Blacklight
- Efforts led to new MARC field to record the work identifier
- Looking to test Work models. Some models may be less good than others, could be bad data, could be bad modeling
- Generated entities feed into an "Entity Ecosystem"
These efforts can add data to improve clustering, using "synthetic authority records.”

Concepts of Expression like language for translation are being really important.

These synthetic records, xR records now exist for 807,450 works; 1,151,773 expressions; <1200 manually created records

Manually built xR record can help create a cluster of related expressions, e.g. translations.

Whole/part, series issues can be partially dealt with these records.

LINKED DATA WORKFLOWS centered on entity data

Reconciler: Connect legacy data with LD entities

Minter: Create and edit entity records

Relator: Create and edit relationships

2018 Linked Data Prototypes, as mentioned earlier, it’s a triplestoreplus, is a system that permits interaction with a triple store, searching, for one thing

Future work: develop editing entities; define and create more relationships; expand MARC-based data mining, series and wholepart modeling

ALCTS/LITA Metadata Standards Committee

Sunday, February 11
1:00 PM - 2:30 PM
Minutes to those with access to ALA Connect: [http://connect.ala.org/node/273377](http://connect.ala.org/node/273377)

- Introductions
- Data collection project, looking to study/interview different groups working in different parts of metadata, partly for their general backgrounds, and also to see how they are doing as far as any work they might be doing about dealing with metadata for underrepresented groups.
- Most of the groups: NISO; DPLA; HathiTrust; VIAF; SAA; NOAH; VRA; IMLS, not pursued; Portico; PCC; ALA AITP; IFLA; Share; ARL; CCM; OLAC; MARCIVE; ...others. MLA was not on the list and I asked for us to be included to be surveyed, as MLA has done some significant work in this area.
- The members are taking on ca. 2 per person and getting to them as possible.
- The hope is to have some good content for a session the ALA New Orleans meeting (Summer, 2018).
- Some “aha” moments from Erik Mitchell. National Library Service: Included an issue with a lack in the MARC record about ways to address accessibility metadata. PCC: had some specific groups working on language/vocabulary issues as far as how they relate to marginalized groups; issues of BIBFRAME and the possibility of decentralized authority control, liking the idea that communities could take control over how they are described. VIAF would like to see more harmonization of how RDF is used; struggle between mandating a standard and giving people freedom. DPLA’s MAP: they’re an aggregator, but they don’t own what they make available. Curious about how they might enhance their processes that might feed into the goals. CCM mentioned that numerical features of MARC were more neutral than English-language focused ways to encode metadata, e.g. labels for ontology properties.
- Worked on the presentation details for the ALA presentation. "How Metadata Enables or Inhibits Discovery for Diverse Communities and Concepts." Title was adjusted with modifications suggested by ALA. Also thought about ways to publish some of the findings of this project. Discussed format for the presentation, wanted to enhance audience participation.
- Last agenda item was to review goals/work plan. Talked about discussions with other organizations, liaisons, some which has been somewhat dormant but could be enhanced. Discussion about tools to use when conducting committee business. Use LITA site? ALA Connect in a re-launched version when it goes live in April.

ALCTS CAMMS Heads of Cataloging Departments Interest Group

Monday, February 12
8:30 AM - 10:00 AM
Program on ALA Connect: [http://connect.ala.org/node/272256](http://connect.ala.org/node/272256)

Next Steps for FAST
Kate Harcourt, Columbia

- Investigated OCLC’s implementation of FAST in WorldCat in order to understand OCLC’s reasons for developing FAST, articulate value and usefulness related to FAST, understand OCLC’s business model, identify obstacles to continuous development
- Started by documenting use cases
- Did conference calls, Columbia/Cornell folks plus OCLC people, learned that FAST wasn’t originally envisioned for large libraries, though its faceted nature is better for linked data than topical strings
- Developed understanding
- Generated a survey, look at usage/uptake/barriers/enhancements to user experience
- Initial results to 14 question survey, 586 individual responses
- More use of FAST than anticipated
- Drawbacks to use may be changing, e.g. topic strings are used more and systems may not support FAST
- Perceived benefits mirror “FAST 5’s use cases
- Some critical features ID’ed: ongoing maintenance, easy Fast lookup, easier integration into cataloging workflows
- Next steps: OCLC says Version 1 in production out this summer, developments will hinge on uptake / Look at more use cases (OCLC with large and British Library)
- More analysis ahead, but fairly detailed results are forthcoming this week

Embedding OCLC Work Identifiers: and Example Workflow
Jackie Shieh, George Washington Libraries

- GW Works Project, includes:
  - BIBFRAME experimentation
  - URI in $0
  - Authority control—authority control had take a back seat because of keyword searching, but adding links brought back the value of authority control
- Looked to integrate OCLC work identifiers. You can reveal OCLC works in WorldCat
- At first didn’t have a place in MARC to hold a WorldCat identifier, and added it manually in 787 $a OCLC work ID $o [#]
- Terry Reese at first did screen crawls to harvest the identifiers.
- MARC Proposal 2017-09 defined field 758, Resource Identifier
- OCLC Work IDs: Highly experimental, frozen since 12/2015, algorithm refinement continuing, IDs may change, if concepts change OCLC may do redirects from one ID to a more modern one, can be harvested via CURL or even web query
- GW has 1,010,047 bib items, 1,010,047 work IDs, 823,895 FAST headings / can’t add more IDs because OCLC has frozen the project
- GW would really like OCLC to keep going
- VIAF pulls identifiers out of MusicBrainz
- Bibliography on slide with good resources to get started
- SHARE-VDE and BIBFRAME —Linked Data in Real Life

Philip Schreur, Eric Mitchell, Michele Casalini

2011 meeting at Stanford on linked data started many discussions, but still using MARC 7 years later....

Some discussion about MARC, how it started in the 1960s, MARC does things well but it’s limited
Reasons to leave MARC: It started out to re-present catalog card data, little semantic meaning in the structure. Problems with internal linkages in compilations.

Why linked data: We can join our patrons who are going to the web for information, linked data’s semantic structure can interact better with the web, can use tools like Blacklight to better synthesize and present data, take advantage of international library data, we can make the move to the web

SHARE-VDE: focused on cataloging workflows based on linked data

Casalini developed SHARE-VDE Project, key element with shared community data

Berkeley and the SHARE-VDE project

Working on Phase 3 of the project, ca 10 libraries are involved

Open community-scale, conversion, aggregation, access, working at the computation scale, not the human scale, ending barriers to adoption

Phase 1 started with small piles of records as a proof of concept

Phase 2 was doing the above at scale, included computer aggregation

Phase 3 required technical developments, developed use cases:

Publish all data sets on the platform, edit them, and have some tools to manage the dataset / SHARE-VDE should support batch or automated updating / support dissemination of data to contributing libraries automated, batch / Libraries should be able to employ the data in SHARE-VDE / Share-VDE should allow us to move to network-level editing (complex, metadata provenance issues?) / Support library needs, somewhat an undefined point / support the ability to do cataloging activities using 3rd part tools

Michele Casalini:

ALIADA predecessor

Phase 1 of SHARE_VDE included data created under different rules

Phase 2 included enrichment of MARC records with URIs

Heading towards embedding multiple ontologies, not necessarily BIBFRAME

Displayed workflow of the components of the processes involved in conversions, including RDF conversion

Identity identification: Defining on the resource and the associated entities, pulling from VIAF ; automated reconciliation is important for conversion, manual for original description / Authify—service for “Relator term detection”

Manual workflows: Includes URI management system, with local library profiles, looking at issues of provenance

RDF conversion includes ALIADA tool

Michele Casalini:

ALIADA predecessor

Phase 1 of SHARE_VDE included data created under different rules

Phase 2 included enrichment of MARC records with URIs

Heading towards embedding multiple ontologies, not necessarily BIBFRAME

Displayed workflow of the components of the processes involved in conversions, including RDF conversion

Identity identification: Defining on the resource and the associated entities, pulling from VIAF ; automated reconciliation is important for conversion, manual for original description / Authify—service for “Relator term detection”

Manual workflows: Includes URI management system, with local library profiles, looking at issues of provenance

RDF conversion includes ALIADA tool

Some questions from the session:

$0s are keyword indexed in Voyager; $0 in Sierra is not indexed
Is SHARE-VDE in a state where data can be extracted? Phase 1 results, yes; Phase 2, no, more difficult to achieve.

Some questions from the session:

$0s are keyword indexed in Voyager; $0 in Sierra is not indexed

Is SHARE-VDE in a state where data can be extracted? Phase 1 results, yes; Phase 2, no, more difficult to achieve.

Issues with metadata provenance and priority

Look at share sharevde.org

Philipp: LD4P2 starting later this year, hope to include SHARE-VDE work in this phase

How far are we from 3rd party tools that can be used for workflows: Not there yet; LC and CEDAR are 1st-gen tools that maybe someone will improve before too longer; PCC Sandbox could be a nice way to develop editors

ALCTS Forum: The Case for Making Video Content Accessible
Monday, February 12
10:30 AM - 11:30 AM

Villains Iglesias, GVPI, Director Business Development

- Speaker is with a tech firm that has the product ElementsPlay, a streaming media component to help media be more accessible online.
- Started with a survey of institutions as far as what they wanted, accessibility was a big item. GVPI trying to get publishers to supply more accessible content. They are also working on how to make this work more discoverable.
- Accessibility has many components: Closed captions (absolute minimum), accessible player (must support many needs such as screen reading, keyboard shortcuts, transcripts, audio descriptions
- Reasons to make accessible: Legal mandates, wide audience includes 5% of the population with hearing loss, user experience (e.g., closed captions to help screen out background as well as helping people understand on-screen language), increased user engagement
- Open web discovery (Closed-caption items are higher in a Google search result), full text will be searchable, reusable content if marked up semantically
- Hurdles: heavy on tech
- Upside: far-reaching benefits, ROI
- ElementsPlay: video solution for publishers
- Has video with/without closed captions, transcripts, ways to move the content around on screen.

Danielle Whren Johnson, Loyola University
Accessibility and Video Captioning at Loyola Norte Dame Library

- Discussed campus decision-making to go towards making things more accessible, both carrot (legal) and stick (benefits to the whole person, not truly educating the student)
- They have hosted videos, expect that 3rd party content has accessible features
- Issues with library-hosted content, looking to see if creators have things like script
- Library create content: working for that to be accessible
- Campus created content: often need to add accessible content
- Captioning tool: Amara
- Original version is open access, so the community can view and contribute content
- Hosted content that integrates Ensemble Video with Amara. They pay for captioning services. $1/hour for transcripts from YouTube automatic captioning.
Preservation and Accessibility in Library Media Collections
Stefan Elnabli, UC San Diego

- We have an ethical mandate to make content accessible? How does it overlap with media collections?
- Talking about a project to reformat media titles and the accessibility concerns in a digital repository. Landscape: Office for Students with Disabilities, Library Digital User Services Program; has media in multiple formats; showed Kanopy resource which has accessibility features; digital media reserves, JW player used for the workflow and has a way to accept uploaded files for captioning. Digital collections: currently can deliver complex objects (e.g. file plus a transcript) but without syncing one resource to the other.
- VHS Migration Project: A way to reformat VHS-carrier materials. Replaces VHS with DVD when purchasable (90%), replaced with similar versions of same resource (5%) in-house digitizing of remainder (5%). Section 108(c) of the Copyright Act describes situations when we can make copies.
- VHS Migration Project included VHS with CC, that we’d like to port over. The conversion process has an uncompressed master that needs to be transformed into CC on DVDs. Working with vendor who will deliver digitized content with a caption file.

QUESTIONS: How do we get buy-in? Library works with accessibility office. It’s important to get IT on board.

Does anyone have experience generating transcript files from extant transcripts? There are ways to add time syncing into a text.

RDA Linked Data Forum
Monday, February 12
1:00 PM  2:30 PM
(This session will be recorded and made available online)

nonRDA is the new nonMARC : Bibliotek-o and RDA
Steven Folsom, Cornell

- Bibliotek-o is a way to look at BIBFRAME functionality and look at alternatives
- Bibliotek-o will hope port into BIBFRAME
- BF is limited in workwork relationships / Bibliotek o reuses RDAu to fill in that gap, look at Relations Pattern document for more detail
- B:Activity: roles as activity? An established ontology design pattern / loose adherence to RDA
- Content and Carrier subclasses of bf:Work / bf:Instance
- As we move out to non RDA data sources we need to worry about what benefits from RDA are to be looked out for
- Rob Sanderson LOUD Linked Open Usable Data, looking at the range of tools versus richness and usability.
- We’re colored by expressing RDA within MARC. We can look at areas that still need to be enriched in RDA when we try to express things.
- No model is perfect. Hope to bridge MARC and nonMARC with RDA.

RDA and Linked Data
Kathy Glennan, ALA Representative to NARDAC, University of Maryland

- Background: RDA is a bundle of standards and use guides, not just a content standard
- RDA and linked data goes back to 2007, formed DCMI/RDA Task Group
- First RDA vocab OMR in 2011
- RDA Registry launched 2014
- RDA moving towards a data dictionary
- Different pieces for different types of users: RDA Reference, Toolkit, etc.
- RDA Reference stored in OMR in RDF linked data format, primary source o RDA Toolkit, includes translations
- OMR Example — Form of Work
- RDA Vocabularies is RDA Reference export to a Git Hub repository
- RDA Registry provides links to download the individual element sets in the current release of RDA Vocabulary’s
- RDA Vocabulary Server
- RDA Toolkit bundle of RDA
- RIMFF visualization and cataloging and prototyping tool to show what is possible
- RDA Implementation Scenarios
- Flat file, card use to Linked authorized access points

James Hennelly, RDA Toolkit, for Kate James, RDA Examples Editor
- Examples in RDA Toolkit with allow displays
- Basic set (as now)
- Recording methods
- View in Context
- View as Relationship, a linked data visualization
- The linked data visualization is a graphic representation of relationships
- DITA stores the information that is visualized

Questions:

Kathy talked a bit more about RIMFF capabilities, a great way to train on RDA but is limited for other possible purposes