ALA ANNUAL CONFERENCE REPORT

New Orleans, LA, June 22-25, 2018

Reports from

Reported by: Rebecca Belford (Oberlin College), Chair, MLA CMC Vocabularies Subcommittee

ALCTS CaMMS Subject Analysis Committee (SAC) Business Meetings

June 24 and 25, 2018

In addition to formal member and liaison updates, the two-part SAC business meeting included a few new business items and updates. The SAC RDA Subcommittee was formally disbanded, based on the announcement by the RDA Steering Committee (RSA) that RDA will not address subjects, as was originally planned. The PCC will revive the practice of having a designated person report from the SACO/PCC at large meeting to SAC. It was decided that the liaison should be from the SACO program in general, not just the ALA meeting. Paul Frank (LC) volunteered to fill this role. SAC received a response to a request for comment on LCSH "Illegal aliens"; ALA will not be pursuing at this time. It was noted that the Sears heading was changed to "Unauthorized immigrants" and that sears is a controlled vocabulary available for subject usage. Finally, per LC Policy and Standards Division, **LC's moratorium on LCDGT proposals is still in place.**

Member and liaison reports, excerpted:

Report on the Sears List of Subject Headings (Maria Hugger)

The *Sears List of Subject Headings* was updated through June 2018. All additions and revisions were routinely updated in the Sears EBSCOhost database and were also available to electronic users as MARC authority files. All headings are RDA compliant. During 2017, 44 new headings were created, 485 base headings were revised, and 44 extended headings were revised. All headings using the word "Elderly" were revised to "Older people" as a heading that presents a broader range of people without the connotation of frailty.

Report from Library of Congress Policy and Standards Division (Janis Young)
Full report available at LC website here.

BIBFRAME (Bibliographic Framework Initiative). The Network Development and MARC Standards Office (NDMSO) provisionally agreed to host the ALA ACRL RBMS (Rare Books and Manuscripts Section)

vocabularies at http://id.loc.gov and began transformation of their new unified structure (still in development). Staff in NDMSO continued to meet with LD4P (Linked Data for Production, now called LD4 or Linked Data for All) partners (Stanford, Cornell, Harvard, Columbia, Princeton and Iowa), to exchange ideas about BIBFRAME databases, triple stores, editors for BIBFRAME RDF, etc. NDMSO began experimenting with use of the Performed Music Ontology that was developed by a project led by Stanford under the umbrella of the LD4 grant.

New Editions of LCSH, LCGFT, LCDGT, LCMPT, and LCC. The 40th edition of *LC Subject Headings* was published online in PDF format in February 2018. The 2018 editions of *LC Genre/Form Terms*, *LC Demographic Group Terms*, and the *LC Medium of Performance Thesaurus for Music* were published online in PDF format in February 2018. The 2018 editions of Library of Congress Classification schedules and tables were published online in PDF format in March 2018. The files may be freely downloaded from the Cataloging and Acquisitions website at http://www.loc.gov/aba/.

"Multiple" Subdivisions. The better to support linked-data initiatives, the Library's Policy and Standards Division in the ABA Directorate will cancel "multiple" subdivisions from Library of Congress Subject Headings (LCSH) beginning in fall 2018. "Multiple" subdivisions are a special type of subdivision that automatically gives free-floating status to analogous subdivisions used under the same heading. In the LC Liaison's Report to SAC, ALA Annual 2018 7 example Computers—Religious aspects—Buddhism, [Christianity, etc.], the multiple subdivision is — Buddhism, [Christianity, etc.]. Over 2,200 multiple subdivisions are established in LCSH, and they can be identified by the presence of square brackets. They generally appear in LCSH itself, as in the heading Computers—Religious aspects—Buddhism, [Christianity, etc.], but some appear in lists of free-floating and pattern subdivisions. The multiples permit catalogers to "fill in the blank" and substitute any word, phrase, or other information that fits the instruction. For example, catalogers can create Computers—Religious aspects—Hinduism because Hinduism is a religion, just as Buddhism and Christianity are. Staff in PSD will create authority records for each valid heading string that was created based on a multiple subdivision and delete the authority record for the multiple subdivision. OCLC will provide files of heading strings that are based on multiples.

As of July 1, 2018, PSD will stop approving proposals for new multiple subdivisions. Instead, catalogers will propose the heading string that is needed. That is, instead of proposing Paleography— Religious aspects—Buddhism, [Christianity, etc.] for a resource about Muslim views on paleography, the cataloger would propose Paleography—Religious aspects—Islam. Proposals that were submitted before July 1, 2018, and that are already under editorial review will be revised to follow the new policy. Catalogers should continue to use existing multiple subdivisions as usual until PSD creates individual subject authority records for each heading string that has been assigned. The multiple subdivision will then be cancelled and catalogers must propose each new use of a subdivision that was formerly authorized by a multiple. Subject Headings Manual instruction sheet H 1090, Multiple Subdivisions, will be revised to reflect the new policy, as will other instruction sheets that refer to multiple subdivisions (e.g., H 1998, Religious Aspects of Topics). The lists of pattern and free-floating subdivisions (instruction sheets H 1095-H 1200) will be revised as the multiple subdivisions are removed from LCSH. Additional details about the project will be announced later in summer 2018.

Short-Term Projects to Revise LCSH and LCC.
American Ethnic Groups and Mass Media. Cutter ranges in P94.5.A-Z, mass media's relation to special groups of people, are subarranged by Table P1, which provides for "General" and "By region or country, A-Z." The classification of materials about the eight American ethnic groups established in P94.5.A-Z has been inconsistent, with some being treated as general works and others being classified by country. In May, the table instruction was removed from the American ethnic groups and the subarrangement was explicitly written into the schedule, with a cutter for the United States.

Intellectual Disability. Proposals to revise the subject heading Mental retardation to Intellectual disability appeared on Tentative List 1806 (June 2018), which will be approved in early July 2018. Approximately 30 authority records for broader and related terms, as well as derivative headings (e.g., Mental retardation

in literature) were revised. Revisions to the classification schedules will appear on a forthcoming tentative list.

LCSH Authority Records that "Duplicate" Name Authority Records. LCSH includes authority records that "duplicate" name authority records when:
- the name heading is used in a general see or general see also reference in LCSH, usually to provide an example for a category of name headings; or
- information pertinent to LCSH assignment must be provided, and that information cannot be provided in the name authority record.

Formerly, examples in general see or general see also references were provided as a matter of course, but the current practice is to avoid doing so whenever possible. In early 2018, all of the "duplicate" records were examined to determine whether they should be retained according to current practice. Approximately ten general see also references were revised to remove examples, so that the "duplicate" subject authority record could also be deleted. A limited number of "duplicate" records were retained in order to preserve scope notes that provide instructions on assignment of the heading as a subject and/or to preserve general see also references. In addition, some records were retained in order to provide UFs for synonymous words or phrases, when those references would not be appropriate for the Name Authority File. Examples of records retained for these reasons include Catholic Church; Qur'an; and Bible. English.

Beat Literature. To describe collections of beat literature or works about beat literature, the heading Beat generation was assigned along with the appropriate headings from the discipline of literature (e.g., American poetry—20th century would be post-coordinated with the heading Beat generation to describe a collection of beat poetry by multiple authors). A series of proposals was made to establish beat as a specific genre of literature and to revise the heading Beat generation to Beats (Persons). The proposals appeared on Tentative List 1806, which will be approved in early July.

Relationship of the Work Being Cataloged to the Classification Proposal. Fully cataloged bib records are not always publicly available in the library's catalog, which makes approval of the proposals much more difficult. For now, the policy of optionally adding a clear statement of the work's relationship to the proposal will remain unchanged, but PSD encourages all catalogers to add such statements to their proposals when the associated bib record either is not publicly available, or does not include the appropriate subject headings. PSD thanks the catalogers who include such statements as a matter of course.

Literary Author Numbers. LC has discontinued the program whereby PCC libraries could suggest LC verified literary author numbers for authors not represented in LC's collections. PCC libraries may instead record locally assigned literary author numbers in name authority records using the coding 053 #4 $a [number] $5 [MARC organization code]. MARC 21 Authority Format field 053 is repeatable; at this time there is no limit to the number of locally assigned numbers that may be present in a name authority record. If LC later assigns a literary author number that agrees with one assigned by another PCC member, the coding of the 053 field will be revised to indicate that the number is now LC-verified. If LC's number does not agree, then an additional 053 field will be added for the LC number; any fields that contain a locally assigned number will be retained in the record.

Genre/Form Terms: Compound Terms for Genres of Literature. Terms that combine two or more existing genre/form terms are generally prohibited from LCGFT. In March 2018, the terms Gay erotic comics; Gay erotic drama; Gay erotic fiction; Gay erotic films; Gay erotic poetry; Gay pornographic comics; Gay pornographic films; Lesbian erotic fiction; Lesbian erotic films; and Lesbian erotic poetry were found to be in violation of the policy because each can be covered by post-coordinating two existing headings. For example, Gay erotic comics can be covered by Gay comics and Erotic comics. The above-listed terms were therefore cancelled. Former heading references to the replacement terms were provided for all of the cancelled terms.

Demographic Group terms. Moratorium on Proposals. Library of Congress Demographic Group Terms (LCDGT) is intended to describe the creators of, and contributors to, resources, and also the intended audience of resources. Terms may be assigned in bibliographic records and in authority records for works and expressions. The moratorium on proposals for new and revised terms that was enacted in February 2018 is still in place while PSD thoroughly evaluates LCDGT's structure and principles.

Q & A at the meeting. Q: What about LCDGT proposals that have been submitted but did not appear on a tentative list? A: Will be approved if they remain appropriate after LC has considered structural issues with the thesaurus. Q: Will LC's next-gen system be MARC-based? A: The group investigating is interested in RFIs that address data formats beyond MARC. Though unsure if the system will supersede MARC, they want to know how it would handle various formats. Q: Will "Dictionaries--Language" be included in multiple subdivisions project? A: Those terms appearing as full headings will be easy to deal with. Other "etc." that may be problematic includes character names, such as "Shakespeare, William, ‡d 1564-1616 ‡x Characters ‡x Falstaff, [Margaret of Anjou, etc.]" Staff may also check volcano eruption dates for typos before fully establishing.

Report of the liaison from CC:DA (Robert Maxwell)
3R Project (RDA Restructure and Redesign). RDA is undergoing a significant revision based on the new IFLA Library Reference Model (LRM). The beta version was released June 14, 2018. A fuller rollout is expected in August or September. Two major issues involve changes in practice due to the IFLA Library Reference Model (LRM): the treatment of serials, and of non-human personages. This has not been totally worked out as of the beta rollout.

Since Midwinter CC:DA approved the report of the Faceted Subject Vocabularies Task Force. This report provided feedback to the ALCTS/CaMMS Subject Analysis Committee on the white paper "A Brave New (Faceted) World: Towards Full Implementation of Library of Congress Faceted Vocabularies." Additional updates at business meeting: The beta version of RDA will remain 'beta' to at least to October 2018, with changes during that time. There will be a significant new beta release in September. Stable text is still needed. When beta phase is declared done, RSC will approve, the RSC board must unanimously approve, and then beta will officially be live; the current version will stay for one year. Maxwell recommended Jamie Hennelly's presentation to attendees.

Report of the SAC Research and Presentation Working Group (Brian Cain)
Brief discussion of whether the working group should pursue maintaining a summary/bibliography of articles and writings. If was agreed that if this task were to be untaken, it would need to be kept current. Discussion will continue. The Working Group is seeking new members; membership is not limited to SAC members.

Report of the liaison from the Music Library Association (Rebecca Belford)
LCGFT. The following LCGFT music terms have been established or revised since the 2018 Midwinter Meeting (in order of approval):

Cantorial music	gp2017026151
Italo disco (Music)	gp2017026149
Concert Music posters	gp2018026004 (heading change)
Fantasias (Music)	gp2018026018
Hip-hop lyrics	gp2017026157
Hip-hop (Music)	gp2017026156
Rap lyrics	gp2017026007
Rap (Music)	gp2014027037 (split from hip-hop)
Romances (Music)	gp2016026068
Trip-hop (Music)	gp2014027146 (BT change)
Underground Electronic dance music	gp2014027153 (heading change)

In nomine (Music)   gp2018026024

Additionally, the following music terms appear on Tentative Lists 1804, 1805, 1806:

Bagatelles (Music)	gp2018026022
Canzonas (Instrumental music)	gp2018026012
Glam rock (Music)	gp2018026029
In nomine nomines (Music)	gp2018026024 (heading change)

Proposals related to the following LCGFT music terms were proposed but not approved since the 2018 Midwinter meeting (in order of response):

Biographical operas	gp2018026014 The proposal provided usage of this term, but it did not provide a definition. If a definition can be found in standard music reference or other authoritative sources, the proposal may be resubmitted. If a definition cannot be found, the term Operas should be assigned along with subject headings appropriate to the topic.
Opera excerpts	gp2018026016
Chinese opera excerpts	gp2018026017 The meeting chose not to approve additional narrower terms of Excerpts that are made up of the term followed by the word "excerpts," which would represent only the form. The term Excerpts may be assigned along with another term appropriate to the work being cataloged. Existing narrower terms of Excerpts will be retained exceptionally. The proposals were not approved.
Rock posters	gp2018026005 To prevent a proliferation of terms for posters from every conceivable musical form, the term Music posters (also on this list) should be assigned with an appropriate subject heading for the type of music or event. The proposal was not approved.
Musical fiction	gp2017026120 …the proposal continues to show that "musical fiction" does not have a unified definition. The proposal was not approved.

MLA/CMC Vocabularies Subcommittee (MLA/VS) continues work on several projects to develop areas of LCGFT needing refinement or reexamination. Discussions for the following close on June 1, 2018:

- Improvisations (Music)
- Reductions (Music)

LCMPT

The following LCMPT terms have been established or revised since the 2018 Midwinter Meeting (in order of approval):

OOVE	pp2018015003
sybyzghy	pp2018015002
unison chorus	pp2018015001
Talerschwingen	pp2018015005

Additionally, the following music terms appear on Tentative Lists 1805 and 1806

subcontrabass flute	pp2018015006
handchime choir	pp2017015017
kudyapi	pp2018015008

As of May 2018, SACO is accepting proposals for music LCGFT and LCMPT terms. SACO catalogers may submit proposals through their existing mechanism, but are encouraged to submit proposals through the SACO Music Funnel (coordinated by Nancy Lorimer, nlorimer@stanford.edu) for expert guidance.

Catalogers at institutions that are not established SACO contributors may also submit proposals through the SACO Music Funnel.

OCLC Music Toolkit for generating faceted music data
The following description is primarily quoted from Casey Mullin's MLA/CMC April 20, 2018 blog post announcing the Toolkit availability:

> The MLA/VS is pleased to announce the availability of the OCLC tool for automated generation of faceted terms based on Library of Congress Subject Headings for music and related coded data in MARC bibliographic records. The Music Toolkit, developed by Gary Strawn (Northwestern University) is an OCLC macro that incorporates a program written by Strawn which analyzes existing bibliographic data and generates corresponding faceted terms from the Library of Congress Medium of Performance Thesaurus (LCMPT), Library of Congress Genre/Form Terms (LCGFT), and Library of Congress Demographic Group Terms (LCDGT), as well as other faceted metadata such as dates and geographic place names. The Toolkit output is then evaluated and adjusted as necessary by the cataloger, who is responsible for correct application of LC faceted terms and their corresponding MARC fields…

> This tool streamlines the retrospective enhancement of bibliographic data for music resources, further advancing the eventual goal of full-scale implementation of faceted vocabularies, and the enhanced discovery of music resources that this data enables. The underlying algorithm, developed by the MLA Vocabularies Subcommittee, can also be programmed to perform batch record enhancements on entire bibliographic databases. Thus, the Music Toolkit, in addition to being a time-saving tool for music catalogers, is a testing mechanism for Strawn's program and the MLA algorithm itself. All who work with music bibliographic data are invited to … install the Music Toolkit and provide feedback to MLA/VS accordingly.... Please be aware that the MLA algorithm, as well as Strawn's program and Toolkit, are all works in progress, and MLA/VS and Strawn will continue to collaborate to refine and improve it.

The release includes the technical report, "Retrospective Implementation of Faceted Vocabularies for Music" (Mullin, April 19, 2018, available here), "Deriving 046, 370, 382, 385, 386, 388 and 655 fields in bibliographic records for notated music and musical sound recordings" (Strawn, March 10, 2018, available here), the Music Toolkit itself with accompanying documentation (Strawn, available here), and a feedback form (available here). In turn, Strawn's documentation provides links to the full set of spreadsheets used for mapping LCSH to faceted terms. Prior to the release, Gary Strawn and Casey Mullin presented on the Toolkit on February 2, 2018 at the MLA annual meeting in Portland, Oregon.

NISO Standard Review. The Music Library Association's liaison to NISO (Nara Newcomer) was asked to comment on ISO 25964-2:2013, Information and documentation -- Interoperability with other vocabularies. A change was proposed to section 20.1.5 Types of subject heading schemes, footnote 12, to reflect developments in music-specific vocabularies since the cited *Music Subject Headings: Compiled from Library of Congress Subject Headings*.

Report of the liaison from the Art Libraries Society of North America (ARLIS/NA) (Sherman Clarke)
The Cataloging Advisory Committee (CAC) was thrilled to have the art and visual works genre/form terms on a Tentative List early in 2018. The terms were discussed, some changes were made, and the terms went live not long after Midwinter. A handful of terms were raised by comments from LC or others and CAC has been completing proposals for about six more headings. When those are prepared and sent to LC, we think the project phase of the "art project" will be complete. It is our hope that the area of visual works can then be opened for new proposals as other project areas have been.

Our other major area of discussion continues to be ArtFrame, the ontology extension for BIBFRAME which is part of the LD4P grant. That work is centered at Columbia. The ARTFRAME group has been working closely with the Rare Materials Group, centered at Cornell. The discussions are wide-ranging and include both descriptive as well as subject metadata. One of the issues which is currently under

discussion is Style/Culture which straddles the traditional ALA committee split between descriptive and subject issues.

Report on the Library of Congress CIP Program (Caroline Saccucci)
As of April 2018 for Fiscal Year 2018, the LC CIP Program cataloged 28,843 CIP titles. In addition, members of the ECIP Cataloging Partnership Program cataloged 4,360 CIP records. PrePub Book Link is the name of the cloud-based system that will replace the aging ECIP Traffic Manager. PrePub Book Link will run on a Unicode-compliant ServiceNow platform and will include enhanced functionality, such as the use of MS Word and PDF file formats for galley attachments and a robust MARC editor tool to convert application data to a MARC record and import into Voyager or OCLC. Collaborations with Bowker/ProQuest, the Book Industry Standard Group (BISG), and the ONIX Best Practices Group produced a mapping from the ONIX schema to the CIP application; this will enable a publisher to search an ISBN and have the ONIX prepopulate much of the CIP application. Publishers participating in the Harvard Online Author Questionnaire (OAQ) will be able to include the unique URL to the OAQ entry for that author, and this will provide richer background information about the author; publishers can also include ISNI and ORCID identifiers for authors. PrePub Book Link is scheduled to launch in the fall of 2018.

Report of the Dewey Classification Editorial Policy Committee liaison (Deborah Rose-Lefmann)
Going forward, the committee's in-person meeting with be in the fall (next meeting: October 2018), so full reports will be made at ALA Midwinter in the future. European Dewey is now a full member. Activities include cleanup for groups of people, such as making the distinction within "child rearing" between the groups being cared for and the groups doing the caring. There is a challenge to classify video games by genre; it is acceptable to use literature and media tables.

Report of the Library of Congress Dewey Section liaison (Caroline Saccucci)
As of April 2018 for Fiscal Year 2018, the LC Dewey Program assigned Dewey Decimal Classification (DDC) to 26,073 bibliographic records; this number includes the 9,687 DDC assigned to CIP e-book records. In order to foster better and more open communication between the LC Dewey classifiers and the OCLC Dewey editors, Caroline Saccucci, CIP and Dewey Program Manager, Library of Congress, coordinated with Jody DeRidder, Director of Metadata Frameworks at OCLC, to host monthly meetings to discuss topics related to their work. As a result of these meetings, which began in March 2018, the weekly agenda is sent to the classifiers, who are welcome to attend when an agenda item is of direct interest. These meetings and the invitation to weekly editorial meetings have already proven to be successful. Rebecca Green, Dewey Editorial Program Manager, will retire at the end of June 2018.

Report on Dewey Decimal Classification and OCLC Dewey Services   (Alex Kyrios)
WebDewey enhancements will be coming soon and are active in test. The notifications feature will be robust, and the contribution feature will be turned back on which will allow direct sending.

Slides from the OCLC Dewey Breakfast/Update at ALA Annual expand on many of the items in the Dewey reports and are available from OCLC here.

Update on the FAST project (Jody DeRidder, OCLC)
Recent Developments
- The results from the FAST survey conducted in November and December 2017 by OCLC and the FAST Five group was released in February. In the survey, there was considerable interest in a production version of FAST, conversion services, and integration of FAST into cataloging software.
- In January and February of this year, Marti Heyman and Jody DeRidder interviewed 10 institutions to gain an in-depth understanding of our users' needs. We particularly wanted to know if FAST in production should differ from what it is now. Interviewees included the Brown, Columbia, Cornell, Harvard, Yale, Stanford, Penn State, the British Library, and two non-library institutions in Australia who are using FAST as their primary method of indexing and retrieval.

- The results of our findings were then shared out with the participants in March, and over the course of 3 sessions in a 4-day period, we reviewed and discussed the findings with representatives from those institutions.
- After some discussions, we determined that the primary source for FAST should continue to be LCSH, but that the community clearly needs support for new and alternative terms. New terms would address areas not covered by the content normally cataloged by the Library of Congress, and alternative terms would provide culturally appropriate options for use which are equivalent to the current terms.
- Any new and alternative terms will need community approval through a transparent process, and an editorial committee should be able to oversee the process and step in to make decisions in controversial situations. Such an editorial committee could also make recommendations for improvements, provide feedback on services and community needs, and could also provide guidance to the community on use of FAST. This editorial committee can help ensure that the directions we take with FAST meet community needs, and can help to engage the community in further expansion and development of FAST.
- OCLC is now working with 6 volunteers from these institutions to develop the scope and structure of the editorial committee, which we hope to launch this fall. We are also working to develop a testing site where we can explore the workflows needed for effective community engagement in proposing and voting on new and alternative terms.
- At the same time, OCLC is working to load FAST as an authority file in production.

Tools for Application and Use
- searchFAST. SearchFAST is a user interface for identifying and accessing FAST authority records and retrieving WorldCat records that have FAST headings
- A FAST Changes page provides access to change files published between updates of FAST downloadable files.
- assignFAST - implements autosuggest technology to facilitate the manual selection of FAST headings. The service enables the integration of FAST assignment into other applications.
- FAST as Linked Data - http://id.worldcat.org/fast/
- The FAST data set is available for download at: http://www.oclc.org/research/activities/fast/download.html
- FAST activity page: http://www.oclc.org/research/themes/data-science/fast.html

Q & A at the meeting:
- Q: Relationship between FAST and LCSH? A: People asking for new a term will need to provide provenance and relationship to existing term and perhaps cultural applicability (US/UK/Australia) so people can request, for example, LCSH FAST with UK overlay.
- Q: Difference between new headings because materials are not cataloged for books vs alternatives for existing concept already in LCSH? A: Alternative terms will have equivalency to existing; new terms will likely cover areas that don't exist, particularly MeSH.

Update on MARC Advisory Committee (MAC) (Stephen Hearn)
SAC will have a chance to think about interoperability connections with subfields $0 and $1 defined in MARC authority format field 024. This differentiates a code that points to a record/descriptive metadata and a code that is real world object metadata. The discussion paper is likely to generate further discussion.

Report from the IFLA liaison (George Prager)
The next IFLA Conference will be held in Kuala Lumpur, Malaysia, from August 24-30, 2018. The Subject Analysis and Access Standing Committee will also be holding an open session program during the Congress, titled "Transforming Libraries via Automatic Indexing: The Impact on Metadata Creation, Discovery, and Staffing Decisions".
The Subject Analysis and Access Section publishes the *IFLA Metadata Newsletter* jointly with the Cataloguing and Bibliography sections.

A new working group on Automatic Indexing was formed during the 2017 conference, chaired by Harriet Aagaard, Information Coordinator of the Subject Analysis and Access Section.

The Genre/Form Working Group continues its work under the leadership of co-chairs Ricardo Santos Muñoz, a member of the Cataloguing Section Standing Committee, and Ana Stevanovic, a member of the Subject Analysis and Access Section Standing Committee. It currently has 15 members, about half of whom are interested colleagues not currently serving on an IFLA standing committee. The Working Group has completed its survey of genre/form practices in national libraries, and published the survey, as well as its report on the survey. The IFLA Survey on Genre/Form Practices in National Libraries Report (31 pages) and the IFLA Genre/Form Survey Gizmo Results (101 pages) are available on the Genre/Form Working Group web page, The Working Group has been considering what should be its next project or projects. It will probably focus on one or more of the following: 1. Monitoring and evaluating the use of genre/form in catalogs worldwide and making recommendations for best practices in catalog design; 2. Investigating genre/form positions in models, principles, rules, and other documents, and: 3. Developing a list of resources relating to genre/form initiatives. The Genre/Form Working Group will be holding a meeting during the Kuala Lumpur Conference on Tuesday, August 28, 2018, most likely after the second SA&A standing committee meeting. Additional information about the working group is available from its web page.

As reported at the last SAC meeting, a joint Cataloguing Section-SA&A group has started working on the revision of *Guidelines for authority records and references* and: *Guidelines for subject authority and reference entries.*

Report of the chair of SAC (Jennifer Bromley/Rocki Strader)
Jennifer rotating off as co-chair; Chris Long incoming. Interns Antoinette and Ethan rotating off; Carla Jurgemeyer and Violet Fox incoming. Member Brian Cane rotating off; Carl Petit incoming. Lia Contursi rotating off as AALL liaison and as SSFV chair. RDA subcommittee disbanded.

Other new business: Subject headings for indigenous persons/topics (Brian Stearns). Discussion regarding general interest in and LC's work with headings for Canadian indigenous peoples/topics. Background: In 2015, the Truth and Reconciliation Commission of Canada released their final report, which included 94 "calls to action", in which Libraries and other educational institutions are prominent for the role that they can play in advancing reconciliation. The University of Alberta Libraries established last year a Decolonizing Description Working Group to identify how we can better represent Indigenous peoples and topics through our metadata, and, from that, created a library resident position to start the work of reaching out to communities for consultation.

SAC Presentation
June 25, 2018

"Supporting Digital Humanities and LAM Data Access through Semantic Enrichment," Marcia L. Zeng.
Slides will be available on the SAC section of ALA Connect (accessible by members only?)

The first hour of the second SAC meeting was dedicated to Zeng's presentation.

A working definition of DH places the field is at the intersection of humanities and computing or digital information technology. Zeng stressed that the scope of data in DH is not limited to digital objects. Data, the reinterpretable representation of information, can also be tangible and ranges from cuneiform tables to tweets. Borgman defines data as "representations of observations, objects, or other entities used as

evidence of phenomena for the purpose of research or scholarship.'" The W3C recommends "provide metadata" always.

Data in LAM (Libraries, Archives, Museums) settings can be unstructured, semi-structured, or structured. Wholly unstructured data (documents, cultural artifacts, original objects, etc.) are diverse in type and comprise the largest quantity, yet are the most challenging to process. Structured data is not necessarily digital, and encompasses catalogs, finding aids, bibliographies, indexes, special collections portals, curated research datasets, and Knowledge Organization Systems (KOS). Semi-structured data can be found within structured data, such as non-EAD archival descriptions or TEI machine-readable text files that have no coded meaning

In LAM DH, the trend has been to transform unstructured data into structured and semi-structured data; or, moving from machine readable data to data that is both machine understandable and machine actionable. Accurate data can be used for interlinking, citing, transfer, rights permission, use, and reuse. With DH and big data, one talks of the known, known unknown, and unknown unknown. There is a shift in research methodology to where the starting point is the unknown unknown, rather than a hypothesis supported by research. This shift is a methodology shift, more than a technology shift. The difficult original step in the process model is determining the scope and what data to use.

There are three perspectives in the creation of structured data. 1) Production: documentation, descriptive metadata, administrative metadata, structural and technical metadata. 2) Content: descriptive metadata, of-ness, relationships, knowledge graphs, indexes, embedded markup, ontology. 3) Audiences' receiving interest: use metadata, citations, research use, tagging, tracking what is searched, shared, followed, liked etc. A striking example of the critical importance of provenance supported by administrative data is the iconic photo of the Kent State shooting--versions published in *Time* magazine and widely circulated edited out a pole from the original photograph.

Phases of the semantic enrichment process are analysis, linking, and augmentation. These are detailed in a Europeana report on enrichment and evaluation from 2015. Generally, enrichment tasks improve metadata about an object by adding new statements. They include:
1. Alignment: starts with existing components in a controlled form
   A. Contextualize: typed relationships between resources of different types. Europeana aligns places with GeoNames; agent names with DBpedia, concepts (GEMENT, DBpedia), time period (Semium). Can manually, semi automatically, or automatically. Can be between two objects, places, concepts. Then what kind of relationships? (Broader, narrower, subject, relation, same, no type). Table available showing vocab, URL, rule, type of entity
   B. Massage: my label matched to their URI or my ID to their URI. Europeana aligns with existing vocabularies. Entities expand to subjects, types. Example: MoMA front end for artist: Wikipedia and ULAN link. Doesn't recreate inside, but points out to work already done. Back end small code "sameAs" code bit (very short). Benefit: with ID, results show up in Google first results page just by adding ULAN ID.
   C. Connect to real things: Example: swissbibMARC aligned to DBpedia and VIAF. Automated workflow for high precision value
   D. FAST entities - also uses sameAs - ties to VIAF so don't need to handle within FAST
   E. Also uses foaf:focus - communicates this Wikipedia is about *x*, which is type of information usually excluded in authority records. mapFAST - uses GeoNames data - based on URI connections instead of putting in data itself
IFLA LRM model demonstrates alignment: resources to place, agent, nomen, time span.

2. Expansion: starts with existing metadata components that are in free form
   A. Parsing data in non-controlled form, to turn into access points. Example Weitz et al Mining MARC… using notes in 508, 511 to create access and role designations
   B. Generating structure from semi structured (entity extraction) as in pull out entities from summary/abstract to provide subject access. Method: Open Calais - paste in text and examine

provided subject access points. Provides occurrence count, relevance, can get to RDF file. Other API: Cogito. API key allows no copy paste usage; auto generates CSV file. Can extract from summary, notes, table of contents. **Are machine-generated results trustworthy?

C. Finding aids description. Used sample of 43 record groups from 16 institutions--examples of semi structured data. Finding that entities related to person, agent, place, corporate bodies tend to be more accurate, higher confidence rating. Social tags, terms, subject headings least reliable. Suggestion: use to start for key entities and topics or as final check to see if anything was overlooked. Subject analysis description and identification pretty good but not for interpretation (about ness) of material.

D. Sample of 44 philosophy theses, selected sample within from KentLINK and OhioLINK, 22 each. Used abstracts and introductions to feed to Open Calais. Human MLIS student matched to determine relevance, type o term, availability in vocab. Result: Average 9 tags per abstract; major concepts correctly identified in most cases. Software often overgeneralized (e.g. "philosophy"), sometimes missed major; Abstract more useful than title. Same approach can be used for museum object labels and descriptions.]

3. Ontology-based design: starts with nonstructured data. Special collection Example: linked jazz oral history transcripts. Natural language processing tool used to pull transcript and find name information, align with VIAF, get data from DBpedia; present relationships based on ontology Example: Online Coins of the Roman Empire, no text to analyze but can model in ontology, every material has URI, can see in addition to descriptive a quantitative analysis. Offers visualization of queries on the fly. Example: images, no text, not easy to determine topic. Deep image annotation, with URI and SKOS coded relationships

Semantic annotation is one level deeper than annotation: enriches with context linked to structured knowledge domain, allows results not explicitly related to original search
Mix n match tool - entries from 1000+ catalogs, tools.wmflabs.org….

4. Exposure
   A. Parallel metadata sets - "head" section of websites, machine readability, Google structured data testing tool
   B. Embedded structured data -- use schema.org and format coding, search engine can read structured data
   C. Enhancing metadata's semantic expressivity

In conclusion, you can enhance through controlled vocabularies, you can link and contextualize across silos, and one-to-many uses of LAM data can support digital humanities. To do this you need a good strategy for using which part of what metadata.

SAC Subcommittee on Faceted Vocabularies
June 23, 2018

The SAC Subcommittee on Faceted Vocabularies (SSFV) held a relatively informal meeting, with a round of updates with discussion. Janis Young (LC Policy and Standards Division) joined the subcommittee to discuss best practices for LC Demographic Group Terms (LCDGT). LC's moratorium on LCDGT proposal submission is still in place.

Rosemary Groenwald reported on the **working group on video game genre/form terms** for authority records. OLAC has applied for its own MARC source code for the vocabulary. There are 80 authority records in progress, with recent work consisting of checking citations. MARCIVE will convert these to MARC records, and the file will be available for download from the OLAC website. The set of terms are a

syndetic vocabulary with top, narrower, broader, and related terms. It is expected that the file will be available at the end of August.

They will be published as a file, without URIs. A decision was made not to publish in the Open Metadata Registry--which would provide URIs-- because it was felt most audiovisual catalogers would turn to OLAC's suite of resources first. There is no maintenance mechanism built in, because the intention was to create a closed vocabulary. Revisions may be made if there are glaring errors. The terms are designed to be general (for example, 'sports' but not 'soccer'). Literary warrant was a priority of the working group in creating the terms. There is no definitive plan or timeline from LC Policy and Standards Division (PSD) regarding when these might be fully incorporated into LCGFT. If LC moves forward with video game genre/form terms, OLAC's vocabulary will be one of several possible source vocabularies.

A topic of recent interest within **Art Library Society of North America (ARLIS/NA)** is how to handle reproductions. For example, there is a question of whether 'photographs' could be applied to a book reproducing photographs. Janis Young (LC PSD) recommended applying the term conservatively for now.

Casey Mullin and Adam Schiff gave an update from the **Orbis-Cascade Cataloging Standing Group**. The group is in the midst of an environmental scan of member libraries, asking about vocabularies usage, rationale, and display configurations. The goal is to have a basis for training and best practices and to provide tools to implement vocabs and create a community of practice within the consortium, which has a shared ILS linked to WorldCat. Results will be analyzed during summer 2018, and they will be presenting findings at a consortium-wide meeting in October 2018. Questions and discussion that followed focused on retrospective conversion or application. The consortial environment presents a challenge: work that has been done at UW requires all records to be re-uploaded to OLCC which in turn prompts an overlay to the consortial catalog as master records are updated. Casey noted that the MLA toolkit is intended for eventual use in OCLC but is not at the necessary level of confidence in conversion; Mary Mastraccio and Lia Contursi have presented on a project using MARC codes for batch adds that may be relevant. UW is working on a related project to identify video recording records lacking field 257 (country of production) and to populate these with local 655$z terms. Until more work is done, terms are not displaying in the consortium's public catalog because subject liaisons are concerned about false misses where the data is not present. Western Washington University will be displaying new vocabularies in their instance of Primo in August.

Other updates: The PCC Policy Committee ("PoCo") is planning fuller a response to the white paper "[A Brave New (Faceted) World: Towards Full Implementation of Library of Congress Faceted Vocabularies](#)".

LCDGT best practices/Demographic Group Terms Manual

Janis Young reports that LC is interested in receiving input on the *Manual*, particular in the form of what topics, situations, or policies should be included. LC has a file of corrections to inaccuracies. The Manual has general instruction sheets on application and the term proposal process, and instruction sheets for individual categories. When done it will cover Work and Expression records (authorities and bibliographic).

Background: Initially LCDGT needed to cover creators to accompany LCGFT; the original plan was to use terms for collections. Questions came later about use in individual works and personal name authority records. There is a need for instruction sheet(s) for personal name authority records. Currently, DCMZ1 does not state LCDGT is required for audience/creator/contributor, but it includes an example. Descriptive specialists in PDF would like to have all LCDGT instructions located together, in the Manual. They have been seeing examples of records where usage does not truly fit the vocabulary. Defined as a 'people vocabulary', there are questions about application for nonhuman entities.

There is a **moratorium on LCDGT proposal submissions** still in place; proposals received to date have illustrated structural and organizational problems including the scope of the vocabulary and hierarchies. Proposals have been coming in with narrow definitions; LCDGT was conceived high-level for collocation, and narrow terms create many issues. As an example of structural problem, "School personnel" has "Teachers "and "Music teachers" as narrower terms, but the narrower terms are not intrinsically part of the broader terms. This does not meet the principle of a hierarchical vocabulary. LC will issue an announcement about updates/changes and hopes to have a more formal update by Midwinter. In the meantime, catalogers needing a term not currently found in LCDGT can use a term from LCSH provided it is coded as $2 lcsh.

Prior to the meeting, SSFV had discussed working on national and regional terms, and now wonders how stable they are. In response: structurally, occupations are more problematic; in application, nationality is more problematic. A discussion about systems' handling of hierarchical facet terms followed.

Returning to situations where additional or clarified guidance would be helpful, perhaps in the Manual: Difficulties often arise when a vocabulary intended for use with works and expressions is applied to people. In particular, retrospective application is not a simple decision or practice. Gender terms appeared frequently in an analysis of authority record errors, in part because of the use of spurious sources. Other LCDGT terms in modern use may, when considered for retrospective application, be inappropriate for the culture and time period of person identified. In general--but particularly regarding gender--there are questions about what constitutes justification or self-identification for use of terms in authority record. An example to consider may be Wikipedia's guidelines for living persons. As an alternative to recording comprehensive information in an authority record, one could consider an authority record as a hub in linked data rather than a source (i.e., "just because you can doesn't mean you should).

Recognizing that there may be major changes made to LCDGT, the SSFV tentative plan is to compile and send to PSD requests for general direction, specific suggestions for specific sheets, or places where something is missing (e.g. instruction for use in personal names). PSD does not have a preference for order of categories. A starting point may be the "National/Regional," perhaps beginning with a discussion of that label versus "Demonym/Geographical".

ALCTS CaMMS Faceted Vocabularies Interest Group
July 23, 2018
Summary by Lynn Gates and the presentation slides are available via ALA Connect here. A FAST update was also submitted as a report to SAC.

"FAST Forward … for the Community" (Jody DeRidder, (OCLC, Director, Metadata Frameworks)
A survey on FAST usage and preferences was followed by interviews with ten institutions in the U.S., U.K., and Australia in early 2018. These were followed by group conversations among OCLC and participants to discuss survey results and future options. Needs and concerns that emerged included long-term support for FAST, improvements to current tools such as converters, additional conversion options and stable APIs, continued open access, clarification of the relationship between FAST and LCSH, maintaining the focus on ease of application by noncatalogers, and the desire to propose new or alternative terms.

There are many Issues and challenges surrounding developing terms apart from LCSH. New terms may be desirable in instances where terms are not accepted by SACO, domain specific (such as medical terms), format specific (journal articles, books), or cover time periods appropriate to cultures outside the U.S. (chronological periods not based on U.S. events). Desire for the ability to request alternative terms may result from terms considered offensive, equivalent terms across cultures (public administration/public

service/civil service), and a mismatch between LCSH and current usage ('climactic change' vs. 'climate change'). If FAST changes are developed as an alternative to SACO, there is a risk that it becomes a duplicative or parallel process. Pros and cons of various methods of proposing terminology were discussed. One audience member argued that potential duplication between FAST and LCSH is not a problem, considering current overlap between thesauri such as LCSH and MeSH, and that LCSH is not faceted. Another attendee encouraged the FAST project get in touch with SAC and other groups, explore options through email discussions, and establish channels to get in touch with catalogers. An editorial policy committee is one option, to oversee community engagement, editorial policies and principles, set priorities, advocate for FAST, and facilitate dissemination. Such a committee supports the idea is that this is mostly non-OCLC and a few from OCLC to provide context and insight on feasibility. There was agreement from the audience that the idea of people able to propose terms in some way is attractive.

Q & A from the audience:
- Q: Is there a parallel process regarding advice from community on tools and APIs? A: Yes, they are deciding whether it is in scope of this committee or another working group.
- Q: Committee/voting structure on new terms: who decides what is culturally sensitive for terms? How do we ensure that expertise exists within community to make those determinations? A: Community would do the voting, not the responsibility of committee. They are open to suggestions.

"Batch Authority Searching with Python: a work in progress" (Kelsey George, UNLV (pre-recorded))
The presentation detailed approaches to a project to enhance approximately 3000 retrospective ETD records from ProQuest with minimal standards. One goal was to update the minimal subject headings in field 650_4 supplied by ProQuest and replace them with LCSH and FAST. The idea was originally to create a batch authority search tool using Python.

To begin, they isolated the subject heading values in 650_4 and 690 (Subject Description, used for LCSH); Subject Code, internal to ProQuest, was ignore. The 6xx fields were exported from MarcEdit to a spreadsheet, then imported from tab-delimited text to OpenRefine to deduplicate values. This resulted in 3,123 rows, one per record. After deduplication of term values, the result was 257 unique values. A text file of those values was used for a batch search of subject headings in Connexion, which resulted in … 11 matches, 20 errors, 226 too many to match. A decision was made to prefer impending deadline over creating a nice tool, and manual authority file searches were made to match authorized term to existing terms. Once mapping was completed, records were batch updated and the MARC replaced. For IR metadata, they used the OCLC FAST converter.

Ideas for a tool when a deadline is not imminent including using pyMARC to search loc.gov, experimenting with MarcEdit validate headings tool, and searching Python against the entire downloaded LCSH file. Questions to ask are whether such a tool would be useful long term, what features it would need to meet the needs of its audience, its user context, and how it would improve interoperability.

Floor discussion
- Q: What kind of session would you like at ALA MW and ANN 2019? A: Followup on FAST project.
- Q: What are you doing at your libraries with FAST? A: Finally stopped deleting FAST locally. [Reporter's note: you are not alone!]

Cataloging Norms Interest Group
June 23, 2018

The session was introduced by co-chair Deborah Skinner. Three presentations followed, with questions at the end of the session.

"Where Does This Go? Cataloging Comics" (Alison Bailund, Hallie Clawson, Staci Crouch (in absentia))

A guide to cataloging comics grew out of the presenters' LIS capstone project at the University of Washington. Their project was sponsored by the Comic Book Legal Defense Fund, an advocacy group for librarians and educators who are dealing with challenges to comics. Acknowledging that comics present cataloging challenges for catalogers not user to comics, their research began with a research survey that informed their creation of "Comics Cataloging 101," with the primary goal of increasing discoverability of comics for users. They designed two research surveys, and, after retweets from two prominent comics authors, yielded 445 responses to the librarian survey and 711 from users. Following surveys, 25 librarians participated in follow-up interviews. Their research indicated that users do want to get comics from libraries, because of the high costs of comics, particularly in situations where the primary object is reading (vs. collecting); they are frustrated by inconsistent cataloging and shelving practices. Librarians value comics in collections because they reach additional types of users, including those preferring to access material in visual form, and comics can add research value to academic collections.

Their resulting document, **Comics Cataloging 101,** is meant to support the cataloging of print comics for the cataloger with no comics knowledge. It covers definitions, assistance with unusual volume numbering practices (for example, a series may have two issues labeled volume "one"), subject heading and genre terms, classification and shelving recommendations, and--of course--a resource list.

Definitions provided focus on basics and granular creative roles. Terms defined include: comics ("a visual narrative using pictures and works used to tell a story for any age group in any genre"), manga/manhwa, issue, volume trade, issue, title, subtitle, reboot, big two (Marvel & DC). Creative terms include writer/author, artist/illustrator, penciler/penciller (spelling varies), inker, colorist, letterer, and cover artist. The authors note that 'illustrator' is overly broad for fans, who may follow inkers in that role. As a result they recommend relator terms that actually cover the type of creative work over broader terms from a controlled list.

External resources recommended for us in cataloging include publishers' sites, Wikipedia, Comic Vine, DC and Marvel wikis (particularly useful to trace characters that appear in crossovers), Diamond books, and Book Riot (particularly useful for awards notes and collection development).

Recommendations from the public service end cover cataloging access points, display, and shelving. Users are well served by being able to access full titles, including volume numbers and subtitles, taking this information from title page verso or outside sources if necessary, and all variant titles. Volume "one" should be identified even though it may often not be labeled as such. General notes indicating trade paperbacks or hardbacks provide useful information; trade paperbacks usually have multiple volumes and are re-releases or series reboots. Access to the entire creative team, with relator codes/terms, is useful for general users and even more so for academic institutions with comics research collections. Apply as many genre and subject headings needed to fully describe items. The presenters recommend shelving comics in their own clearly marked section by title including both fiction and nonfiction; if nonfiction is interfiled, comics can be identified by a spine label or color. Classify, when possible, by character/series title for long-running comics. Front-facing displays and seasonal, topical, or movie tie-ins help promote collections; highlighting diverse stories and characters may be easier with comics collections than with other genres.

The presenters noted that this resource guide places online and web comics, children's comics, and other formats out of scope. Their guide will be available from the Comic Book Legal Defense Fund's (http://cbldf.org) Resources page.

Q&A was at the end of the session.
- Q: Have you tried to add relator terms to the vocabulary? A: They hope to soon; new PCC records have "penciler."
- Q: Preference between shelving by title vs. character? A: Recommend either PN6728 then hero or title.
- Q: What about titles of individual volumes for manga? A: Find it and add it, bracket if you need to. They put this information in field 245.

"Forewarned is Forearmed: Prepping Your Cataloging with URIs in MARC" (Jessica Hayden, University of Northern Colorado)

The University of Northern Colorado (UNCO) Libraries contracted with an authority vendor for retrospective cleanup and RDA enhancement of their entire back file of records in spring of 2017. As part of the service, they added URIs in $0 to authorized access points as a free enhancement.

Why add subfield $0? In theory, its presence will make conversion to BIBFRAME more effective, Zepheira recommends its addition, and there is hope that other (non-catalog) systems such the discovery layer or institutional repository will utilize them earlier. In general, libraries can incorporate these into MARC records in several ways: vendor cataloging, MarcEdit's MARCNext feature, manually with id.loc.gov, or with an authority vendor profile.

UNCO uses Sierra ILS with Encore and Summon on the front end. Before loading the full enhanced file back from the vendor, tests with a sample set revealed that the Classic catalog was displaying subfield $0. They changed the display to suppress the subfield from view, then reloaded the backfile.

Once the file was fully loaded, several problems emerged. The subfield displayed in more places that were anticipated; it caused indexing and browsing issues in the Classic catalog (though not Encore); index, browse, and display in the union classic catalog and Encore were affected by the URI's inclusion in pass-through searches from browses. Fixes required correction to the webpub.def file in III, a III ticket to correct encore, and reindexing of UNCO's catalog (not free and a fix that extended to the union catalog.)

Recommendations for other libraries considering enhancement: experiment with a small set of records with MarcEdit or manual addition first; check *all* access point fields for display issues; check indexing rules first (particularly if there are costs associated with reindexing); and weigh the pros and cons, particularly considering whether reindexing to accommodate MARC changes (e.g. subfield $i) is ongoing. URIs are increasing in use, and are addressed in RDA's four-fold path. Also, coding for URIs changed since UNCO's project where subfield $0 is now used to point to records and $1 to point to the thing itself.

"User Tagging in an OPAC: A Quantitative analysis of 7 years of I-Share User Tags" (Brinna Michael, University of Illinois Urbana-Champaign)
As of this writing, slides have not been posted; a related research poster by the author is available via IDEALS, UIUC's IR.

The study began with a prompt from the data librarian at University of Illinois Urbana-Champaign (UIUC): can we use tags to improve the discoverability of datasets? User tags UIUC are present in I-Share, the VuFind Layer discover layer of the CARLI (Consortium of Academic and Research Libraries in Illinois) catalog. Tags have been added since 2010. Although not searchable, they can be used to collocate items once identified. Users must log in to add tags, and tags are visible to all users.

Analysis was performed on a tab delimited file of student-supplied tags added 2010 through March 2017. Institution, record, user, and tag were available for analysis. Tags were a mix of useful and nonsense. Results indicated that 50% or more were from doctoral institutions and 23% from UIUC. Tags were categorized: content description, title words, commentary, creator, course information, object description, and location or call number. They tracked users per record, tags per record, and tags per user. To

conduct a qualitative analysis of UIUC data, they created a report to pull titles, creators, and subjects for each record with tags, then further narrowed to records that did not have subject headings, yielding a set of 1200 records. They then compared the top ten tags for CARLI, UIUC, and the UIUC set without subject headings. By category, tags were 54% content description, 22% title words, 8% user commentary. For the subject-less records, all top ten described manga.

To the question of why users add tags, one motivation may be context for under-described resources (like manga). An additional use appears to be the creation of subcollections for personal or research use. In the future, I-Share may shift to an annotation service, which will provide options to make annotations public or price and which allows for full sentences. An addition future step is to index user content for search and discovery; they are searchable with I-Share's update to VuFind 4.1. Suggestions for library actions include tapping into the knowledge of users, particularly for specialized genres like manga, and to increase user awareness of the intended functionality and capability of tags, perhaps through information literacy efforts.

ALCTS CaMMS Heads of Cataloging Departments Interest Group
June 25, 2018
Abstracts and speaker bios are available via [ALA Connect](#) (login required).

The three-part session was introduced by Martin Knott, interest group co-chair.

"Collective Cataloging: Sharing the Load across a Consortium" (Ellen K.W. Muller)
Big Ten Academic Alliance ([BTAA](#)) library consortium is made up of fourteen university libraries in the Midwest and Mid-Atlantic. The BTAA Cooperative Cataloging Partnership project to share cataloging also included the University of Chicago. The project was a collaborative effort across the cooperative to catalog items for which individual libraries lacked language expertise. Items represented 99 languages and 10,800+ backlog items. The made a list of languages with and without expertise as a preliminary step to cooperative cataloging. The one-year pilot began in April 2014.

Challenges addressed proactively:
- 1:1 relationships were not possible. Desire was for reciprocal relationships where benefit equal, but not possible because there were no direct matches. Instead, each library would send 100-120 titles and receive 100-120 titles, without requiring exact count match for cataloging.
- Deciding on descriptive standard. Libraries wanted full original cataloging and used full BIBCO standard record, but could choose AACR2 or RDA depending on where each institution was in the changeover. Classification: some institutions still using Dewey. They decided to allow creating institutions to class according to their system, and the receiving library would add their class if necessary.
- How do you define language expertise? Agreed to trust self-assessment for ability to create full BSR-compliant record with appropriate and full access points, subject analysis, and scripts according to PCC.

Unexpected challenges:
- Lack of shared, interactive communication tool. They decided on the Google suite.
- Mailing. Best practices created but not uniformly followed, and mailroom practices varied. Recommendation: have conversations with mailroom ahead of time so they know what to expect.
- Tracking and statistics. Needed a lot of metrics and data to assess success in order to determine if this would be a viable solution. Result was an eleven page form, a challenge to get everyone to fill out each time. Problems included inconsistent data entry, hand-estimated time calculations, time spent tracking time, and determining salary data when not all institutions made data available.

- Time spent: 80 working hours to develop tracking mechanisms and analyze the forms.
- Cost of scanning vs. cost of shipping. Determined that scanning costs significantly more expensive than shipping because scanning would be done by librarians and high-level staff. Bound-withs were a particular challenge.

Easy wins:
- Strong leadership: Communication, project management, negotiation.
- Sharing metadata: All records created in OCLC without holdings set then sent with OCLC number.
- Project flexibility: Original pilot was for monographs and serials but led to expansion to maps, rare materials, DVDs, CDs

Conclusions:
- Meaningful cost savings over vended options. For example, pilot cost of original Korean monograph $17.54 vs. vendor cost of $34.00.
- Economies of scale. No hiring or training needed to process languages.
- Greater need than just language expertise--also need format expertise. Still challenge of insufficient.

The partnership was implemented in 2016, after the conclusion of the pilot. Costs and funding were revisited. They agreed on looking at the consortium as a collective where collections and processing occur together. They established a minimum threshold of ten hours per month commitment. Mailing was resolved by utilizing established options for resource sharing (UBorrow). Regular communication was established, with monthly calls and a group email list. Two years in since pilot:
- Two year memorandum of understanding
- 100% added value in satisfaction survey
- 90% indicated that participation hasn't been burdensome; 10% said tolerable burden because of positive outcomes.
- 40+ languages active
- 2100+ titles cataloged
- Standards: BIBCO/CONSER; other formats also up to full standards within community

The project was awarded the ALCTS outstanding collaboration citation awarded at ALA Annual 2018.

Questions and answers:
- Q: Pilot data was lots of data. What were most meaningful data points? A: I loved all of it! The most surprising outcome was how much people did not need help because of the lack of expertise but because they had capacity issues. In other words, they do have expertise but can't handle volume. It was relieving to know results actually led to cost savings.
- Q: Expansion? A: In conversations with Ivy Plus Libraries. Expansion can introduce difficulty in shipping if sending outside each consortium.

"The Adolescent Institutional Repository: Metadata Management Perspectives and Challenges" (Casey A. Mullin)

Western Washington University's institutional repository ("CEDAR") is a collaborative model of digital collection management, including curation, systems, and metadata experts. Metadata created is used in Western's digital collections (in CONTENTdm), the CEDAR IR, and Western Libraries discovery layer (OneSearch/Primo). Ultimate goals in managing metadata were standardized metadata profiles in CEDAR, protecting metadata in future migrations, and--a long-term goal--harmonizing descriptive metadata practices across all three platforms.

The process developed included seven steps:
- Step 1: Floor for metadata object based on Orbis-Cascade Dublin Core best practices developed for their Digital Public Library of America (DPLA) hub participation.
- Step 2: Fill in gaps in metadata profiles in all 60 collections, including early collections made without a data dictionary and later collections with customized data.

- Step 3: Create complete data dictionaries for each collection. Required clarification of meaning of "required element" and determining public display. Thirty spreadsheets detailing data dictionaries for each ("Welcome to my personal Excel hell") were created. (Step 3 concurrent with Step 2.)
- Step 4: Documentation. Problem: giant spreadsheet shows what but not why, when, who. Solution: create a registry of CEDAR collections with name and creation/history in a shared drive. Problem: giant spreadsheet unwieldy for broader audiences and other stakeholders. Solution: Distilled 30 spreadsheets into a single eleven-page Word document.
- Step 5: Discussion. "G7" meetings - periodic meetings with stakeholders. Decisions made: Art and Architecture Thesaurus (AAT) for genre/form vocab, rights statements, update requests.
- Step 6: Master data dictionary, with common set of core fields and "long tail" of others.
- Step 7: Remediate. Batch revision for 'low-hanging fruit'. Still need to solve identity management and authority control.

All seven steps are continuing, mostly concurrently, with progress on all fronts. Lessons learned:
- Involve IR stakeholders early and often
- "Document, document, document"
- Patience

Questions and answers:
- Q: How long to this point? A: 18 months.
- Q: Are you pushing records to Primo? A: Yes, all along, but records not yet mapped to Dublin Core weren't showing up.
- Q: Comments on coordinating different divisions at first? A: Combination of ad hoc conversations with individuals and getting big group together; try to address as much in the moment when possible. Depends on situation.
- Comment: UNLV working on similar project.
- Q: Did bepress request a meeting? A: WWU IR manager in frequent contact with bepress - no strategic conversations formally.

"Toward Solving Legacy metadata Issues and Improving Discoverability in Digital Collections' (David Van Kleeck, Chelsea Dinsmore (in absentia))

The presentation described a pilot project of the University of Florida (UF) Digital Collections to automate metadata generation. UF digital collections contain 900+ individual collections with 545,000 items. Records have varying amounts and type of metadata and a lot of legacy metadata. Content includes published and archival resources, many of which are not represented in the online catalog. An automatic process is becoming essential to enhance existing and generate new metadata at scale. A long-term goal is the Portal of Florida History, which will include much UFDC content. A primary challenge is to identify, aggregate, and present collections coherently.

For the pilot project for automated metadata generation, UF partnered with Access Innovations company. They selected the Theses and Dissertations collections in UF's institutional repository, making 29,000+ items available for the pilot. They wanted to add Florida topics, but could not just include everything with "Florida" in the metadata because they were all produced in Florida. They used Access Innovations' Data Harmony suite, which is project-based and thus customizable per project.

Project steps:
- Extract full text with optical character recognition (OCR)
- Extract metadata
- Test thesauri to determine the best
- Built Florida-specific taxonomy
- JSTOR thesaurus chosen for breadth and depth
- UFDC data exported to XIS (XML Intranet System)
  - Metadata enhanced
  - Records reviewed; can be exported to OCLC when needed.
  - Send to XIS repository where it can be re-exported to UDFC.

- Florida-specific taxonomy of geographic terms. Believe rule based taxonomy can serve a filter during search process in UFDC.

Before, records contained original subject terms from students and sometimes staff-added LCSH. Existing metadata was in METS XML format. They decided how to tag with topic, location, and name. Decided on 7 topical terms, automated identification of words used most often and matched to JSTOR controlled vocabulary. They kept original keywords and also added controlled terms. Process will improve Florida history identification in particular.

The next project selected is the Florida Cattlemen's Association journal, which contains much Florida family history. 618 issues will be digitized and a sample set selected. Text will be run through software for subject terms and then reviewed to see if 7-10 terms per issue make discovery better vs. 3 terms per title of the serial. Testing phase has not begun yet.

Lessons learned:
- OCR results are variable - merged three resulting sets to be able to use text.
- All collections are special - no one size fits all. Many formats and content types so system needs to be flexible. Always do test run.
- Still need experts but they can use their time to weed out terms vs. assigning from scratch. Some ETDs ended up with 45 suggested, so have programmatically limited to 10. Other collections may need more extensive curation or adjustments up or down to number of terms.
- Standardizing METS fields essential for advanced search and discoverability. In future, all topical will map to a single field. So will names of all types (personal, corporate, geographic). Existing terms will be mapped.

Looking ahead
- Apply XIS to all UFDC content.
- Extract existing records for cleaning and enhancement
- Will host system at UF (currently with vendor)
- Develop better assessment plan of impact of taxonomy changes on discoverability

Looking further ahead…
- Will apply to most cataloging at UF including print materials.
- Can send to digital repository, OCLC, Discovery service as appropriate.

Questions and answers:
- Q: Subject access of journal assessment? A: Can get usage stats from software to see click-through and access counts.
- Q: What kind of OCR? Have you considered Google Docs? A: Initially, Access Innovations but had to run several packages because of varying file quality. UF may purchase an additional OCR package for higher grade. Magazine layout text proved very difficult.
- Q: When system returns JSTOR vocab, does it provide confidence level? A: Yes. Still working on rule building that governs which terms returned. Ten most likely terms have so far seemed of good quality. May ultimately decide to trust after more testing so this can be employed at scale.
- Q: Staff review - does result line up with automated ranking? A: Initially, no. After narrowed to limit of 10, on close alignment that they are nearing satisfaction with results.

OCLC Research Update

June 25, 2018

The full video recording of the session and the slide deck are freely available from OCLC,
https://www.oclc.org/research/events/2018/06-25.html

Sharon Streams gave an update on the project "Researching Students' Information Choices, Determining Identity and Judging Credibility in Digital Spaces" (http://oc.lc/rsic). Researchers are in the data analysis stage for two questions: Do STEM students differentiate among different types of digital resources at point of selections? How do STEM students determine the credibility of digital resources? Other OCLC Research updates covered the "Wikipedia+Libraries: Better Together" project, the ongoing webinar series on "Evaluating and Sharing your Library's Impact," an update to the report on voter perception of libraries, "From Awareness to Funding" (http://oc.lc/awareness2018), and the distinguished seminar series (http://oc.lc/dss), which continues in November 2018 with Rosie Stephenson-Goodknight.

Next, Andrew Pace presented "Prototyping a Linked Data Platform for Production Cataloging Workflows." Phase I, a partnership of OLCC, Cornell, and UC Davis, ended in April 2018. Phase II includes 14 libraries, primarily universities, one large public, and the NLM. Phase II aims to develop an entity ecosystem, build a user community to create and curate data in that ecosystem, and provide services to reconcile and explore data. The detailed slides available from OCLC contain software and platform screenshots and list potential use cases.

The final presenter was Karen Smith-Yoshimura, addressing "Representing Translations in Wikibase". The presentations and slides are supplemented by the related post on OCLC's "Hanging Together" blog. Translations present opportunities for developers to present information in the preferred language and script of the user. Seven percent of the written works in WorldCat have been translated into at least one language. Heidegger's *Sein und Zeit*, with 570 identified translations into 33 languages, is being used as a test case for modeling works and translations in Wikibase. Input comes from WorldCat 'gold star' record metadata and Wikidata descriptions. A closing slide offered a glimpse of the potential in discovery, with an image of the discovery layer for the Wikibase prototype that also brings in external information from Wikimedia commons and DBpedia. Next steps include an automatic classifier to determine if a record represents a translation or not, bulk corrections, and bulk harvesting of Wikidata creative works. An audience member asking about which non-roman scripts can be used; everything is Unicode so there are no restrictions.

LITA Top Technology Trends
June 24, 2018

Each year, for the LITA Top Technology Trends, each panelist selects a current or emerging trend they see as having a current or future impact on libraries and how we might apply those trends to services. Marshall Breeding moderated the discussion with five panelists.

Jason Bengston: Quantum computing, probably realized in six to seven years; the impact will be greater ease in breaking public key encryption easier for symmetrically encrypted data.

Laura Cole: Public housing and technology, looking at ways in which libraries are expanding their digital presences to public house, embedding the public library in public housing.

Justin de la Cruz: Psychometrics, micro demographics and hyper-specific data collection that can be used to change beliefs. Libraries need to be proactive in how we plan execute gathering of data, and how we use data from our users. Libraries should look at what is collected by third parties (for example, libraries

are sending search data to Amazon when Amazon provides cover images) and if we can offer opt-out, what tracking users come in with themselves (leaving their personal accounts logged in on public computers), and the data we gather ourselves.

Marydee Ojala: Death of transparency. "How can we know when results are neutral? Actually, we can't." Librarians need to teach that answers can be affected by machine learning, and that machine learning can amplify bias. Floor comments: We are in a constant state of A-B testing for searches, while the source code is unavailable. We are in a new stage of 'information is fluid'.

Reina Williams: Next-gen Learning Management Systems. Organizational skills required to maximize an LMS can present challenge for professors. Newer trends are for students to build their own sites and spaces, such as Domain of One's Own. Moving forward, what can be offered at the K-12 level?

Questions and remarks from attendees followed; answers come from panelists and the audience.
- Q: What tech from five years ago didn't end up 'all that'? A: Augmented reality; Google glasses; QR codes.
- Q: Take-home tech and ideas of future trends for nontraditional tech to lend? A: Usage within a library could be something like Echo as an intercom feature to fill in places where there isn't a physically staffed service point. A: Google Home action customized for library with a JSON feed for event software so people can ask what's going on (working on tying in to Polaris API). A: Offer login for remote computer with significant computing power, so youth and gamers can access high end computer without actual hardware infrastructure; provide as subscription.
- Q: What does it mean to place tech into homes with implications for privacy?
- Comment: Thinking about student data analytics and privacy/big data issues. On the upside, one can tie student outcomes to library usage for use in budget scenarios. On the other hand, it is not necessarily ethical to capture granular data.
- Q: More on libraries and public health issues? MO: Public library currently looking at ways to offer public health portal to facilitate communication with physicians. RW: NLM has an initiative called All of Us Research Program, collecting unique patient data for precision medicine, which looks at an individual patient in context of similar patients because broad pool of data generalized. NLM is working with public librarians on teaching patrons to help with PubMed, Medline, etc. Studies on information sharing behavior could inform public awareness efforts. Further wrinkle: data security and at what point do rare conditions mean privacy is impossible with condition plus one or two other data points
- Q: Security on cellular network vs wifi? A: Cellular networks hand you off to wireless whenever possible anyway; consider bandwidth limits.
- Comment: Technology remains neutral, but there are many legal and social issues related to tech: "There will always be ways to misuse technology."