Subject Analysis Committee

Presentation: “Pre-Coordinate vs. Post-Coordinate Subject Access: Pros and Cons and a Real Life Experience”

Peter Fletcher (UCLA) discussed the relative merits of pre- and post-coordinate subject systems. Library of Congress Subject Headings (LCSH) is the prime example of a pre-coordinate system, while PRECIS as an example of a post-coordinate system. Among the arguments in favor of pre-coordination are: sophisticated context, which is needed for disambiguation, suggestibility and precision (Elaine Svenonius); and, a stable, comprehensive, globally-used vocabulary (in the case of LCSH). Among the arguments in favor of post-coordination are: more structurally sound; more amenable for Linked Open Data uses; and, inconsistent syntax rules cause difficulty for clean, faceted interfaces (Kelley McGrath).

Although LCSH is the dominant subject system in many libraries, it faces serious limitations for precision, recall, understanding and relevance ranking. And its sustainability is in question if discovery systems (and thereby users) are not taking full advantage of its benefits. Fletcher closed with the prediction (as previously stated by F.H. Ayres) that post-coordination is the future that catalogers must accept, and that subject searching processes should be user-driven.

Diane Boehr (National Library of Medicine) described a transformation in subject access that has occurred at the NLM in recent years. Until 1999, MeSH (Medical Subject Headings) were used just like LCSH, with subdivisions being employed as appropriate. Beginning in 1999, in an effort to harmonize cataloging and indexing practice, NLM began encoding MeSH terms in a post-coordinated fashion; main headings and their subdivisions are now given in separate fields (much like FAST headings are today). This change resulted in a discrepancy between how NLM stored its data internally, and how it encoded its data for distribution (through OCLC WorldCat, etc.). Subject components in the native NLM records would be reconstructed into pre-coordinated headings for distribution. This satisfied the needs of the WorldCat community, but the imperfect algorithm would occasionally result in problematic output (e.g., Eskimos $z Hawaii).
In 2005, libraries who utilize NLM’s MeSH data were surveyed for their opinions on this process. 75% of respondents preferred to “unstring” the MeSH headings in their local catalog anyway. Arguments for and against the discontinuation of the reconstruction process were passionately given, but ultimately NLM decided to begin distributing their records in 2016 with MeSH headings exactly as they store them internally. Other implementers of MeSH are expected to follow suit in WorldCat and in PCC records.

Report of the liaison from the Library of Congress Policy and Standards Division (Janis Young)

LCSH updates:

Online Training for LCSH

In cooperation with the Simmons College School of Library and Information Science, PSD is developing free online training in Library of Congress Subject Headings. The training is being developed primarily to meet internal training needs of the Library of Congress, but it is also being made freely available through the Cataloger’s Learning Workshop as a service to the library community. Training units are divided into two or more modules, each of which consists of a lecture and one or more exercises or quizzes. Technology requirements include an Internet connection and the ability to play audio and video files. The initial modules have been mounted on the CLW at https://www.loc.gov/catworkshop/LCSH; additional modules will be added as they are completed.

Proposal to Change the LC Subject Headings Aliens and Illegal aliens

In response to requests from constituents who consider the phrase illegal aliens to be pejorative and disappearing from common use, the Policy and Standards Division of the Library of Congress, which maintains Library of Congress Subject Headings, has proposed that the headings Aliens and Illegal aliens both be replaced.

If approved, the heading Aliens will be replaced by Noncitizens, which is currently a Used For (UF) reference to Aliens. Illegal aliens will be replaced by two headings: Noncitizens and Unauthorized immigration. Other headings that include the word aliens or the phrase illegal aliens (e.g., Church work with aliens; Children of illegal aliens) will also be revised.

Proposals to revise Aliens, Illegal aliens, and all of the related headings appear on Tentative List 1606a, which may be accessed at https://classificationweb.net/tentative-subjects/1606a.html.

The Library of Congress accepted comments from the library community and the general public through July 20, 2016. Because of the high volume of comments that is expected, comments will be accepted only through an online survey, the link to which is available at the top of Tentative List 1606a. Review of the comments by the Policy and Standards
Division began after July 20, 2016. Final disposition of the proposals will be announced later this year.

These proposals have generated interest in the United States Congress. The FY2017 House Legislative Branch Appropriations Act (H.R. 5325) report includes the following instruction: To the extent practicable, the Committee instructs the Library to maintain certain subject headings that reflect terminology used in title 8, United States Code. The terms "illegal" and "alien" are frequently used together in title 8.

The issue was a key point of debate in consideration of the fiscal 2017 appropriations bill in the House. Several Members suggested amendments, raised points of order, and offered other motions to remove the language or prevent the legislation from moving forward. None of these suggestions were accepted by the House, and the bill passed on June 10.

Separate legislation has been introduced by Rep. Diane Black of Tennessee to retain the headings in the current form. H.R.4926, the Stopping Partisan Policy at the Library of Congress Act, states, “The Librarian of Congress shall retain the headings ‘Aliens’ and ‘illegal aliens’, as well as related headings, in the Library of Congress Subject Headings in the same manner as the headings were in effect during 2015.” The bill has been referred to the Committee on House Administration.

**LCGFT and LCDGT Updates:**

*Genre/Form and Demographic Group Term Manuals*

In January 2016, PSD published drafts of the Demographic Group Terms Manual and the Genre/Form Terms Manual, which are available at [http://www.loc.gov/aba/publications/FreeLCGFT/freelcgft.html](http://www.loc.gov/aba/publications/FreeLCGFT/freelcgft.html) and [http://www.loc.gov/aba/publications/FreeLCDGT/freelcdgt.html](http://www.loc.gov/aba/publications/FreeLCDGT/freelcdgt.html), respectively. The comment period on the manuals closed on May 31, and revisions are ongoing.

*Demonyms for Local Places*

In fall 2015, PSD decided in principle that demonyms for the residents of local places (e.g., counties, cities, city sections) may be included in LCDGT, but the appropriate level of disambiguation among demonyms that are, or that may be, used to refer to people from unrelated places had to be determined, and the form of the qualifier also had to be decided. PSD received several comments on its November 2016 paper entitled “Demonyms for Local Places in LC Demographic Group Terms: Analysis of the Issues” ([http://www.loc.gov/catdir/cps0/lcdgt-demonyms.pdf](http://www.loc.gov/catdir/cps0/lcdgt-demonyms.pdf)), in which several options for disambiguation are discussed. Review of the comments is ongoing. PSD will announce a decision about demonyms for the residents of local places when it is available.

**LCDGT Pilot Phase 3**
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.

PSD has determined that phase 3 of the pilot will continue through the end of 2016. Proposals for terms that are needed in new cataloging only are being accepted. Due to PSD staffing and workload considerations, proposals that appear to be made as part of retrospective projects, or projects to establish terms that are not needed for current cataloging, will not be considered. All proposals should follow the guidelines on form of authorized term, references, scope notes, research, etc., presented in the draft Demographic Group Terms Manual.

SACO members should use the Proposal System when making proposals and send an email to saco@loc.gov to inform Coop staff that the proposals are ready, according to the normal procedure.

PSD is also continuing to accept proposals from catalogers who do not work at LC or in a SACO institution. They may contribute through a survey available at http://www.surveymonkey.com/r/LCDGTproposals. The survey requests the same information that the Proposal System does, but in a simplified format. PSD staff will make the formal proposals, which will be vetted during the standard editorial process. The survey will be available for the duration of Phase 3 of the pilot.

*Report from the SAC Working Group on the LCSH “Illegal aliens”*

After a lively discussion at ALA Annual, the report of this working group, which was incomplete at the time, was submitted for a vote on July 13, 2016. The summary of their recommendation, in response to LC’s proposed change (see above) is as follows:

“This report concurs with the Library of Congress decision to change the subject heading *Aliens* to *Noncitizens*, but recommends that *Illegal aliens* be replaced with *Undocumented immigrants* where appropriate. In cases where the subject heading *Illegal aliens* has been assigned to works about nonimmigrants, more specific terms should be assigned. “

The report will be submitted to Library of Congress through their survey mechanism for this proposal. The full report is available at: [http://connect.ala.org/node/255185](http://connect.ala.org/node/255185)
SAC Subcommittee on Genre/Form Implementation

Summary of Work Done since ALA Midwinter

The SAC Subcommittee on Genre/Form Implementation (SGFI) and its LCGFT for Literature Working Group had a relatively quiet period of activity since ALA Midwinter 2016, but it accomplished a number of notable things:

- Drafted MARC Discussion Paper 2016-DP29: **Defining New Subfields $i, $3, and $4 in Field 370 of the MARC 21 Bibliographic and Authority Formats**. Available at: https://www.loc.gov/marc/mac/2016/2016-dp29.html. The paper was approved by SAC and was considered by the MARC Advisory Committee at ALA Annual 2016.

- Drafted MARC Discussion Paper 2016-DP30: **Defining New Subfields $i and $4 in Field 386 of the MARC 21 Bibliographic and Authority Formats**. Available at: https://www.loc.gov/marc/mac/2016/2016-dp30.html. The paper was approved by SAC and was considered by the MARC Advisory Committee at ALA Annual 2016.

- LC Policy and Standards Division requested that the Literature Working Group provide warrant/usage for the proposed LCGFT term **Stage adaptations** before it is put on a monthly list. Literary warrant and usage for the term and a number of variants was provided, and a revised form of the term, **Theatrical adaptations**, was approved on monthly list 2016-3.

- In light of LC’s announcement in December 2015 of a revised definition of genre/form that will be used for LCGFT (which was a response to a proposal from the SGFI’s Working Group on the Definition and Scope of Genre/Form for LCGFT), the LCGFT for Literature Working Group held discussions at ALA Midwinter 2016 and then over email regarding literature genre/form terms that had either been deferred or rejected from the initial group of approved terms and whether some of them should now be eligible for inclusion. The group decided to ask LC to reconsider thirteen terms, and in April 2016 provided LC with draft authority records and justification for each of the terms (http://connect.ala.org/node/254562). The requested terms are: **Academic fiction**, **Autograph verse**, **Closet drama**, **Complaints (Poetry)**, **Domestic tragedies**, **Fairy plays**, **Fairy poetry**, **Film tie-in fiction**, **Funny animal comics**, **Game tie-in fiction**, **Gentle reads**, **Imaginary voyages**, and **Toy theaters**. As of the writing of this report, there has been no response from LC.

- The LCGFT for Literature Working Group provided feedback to LC on twenty-two ethnic/folk drama terms that were included on monthly list 2016-05. The terms had been deferred from the initial approved list of literature terms.

The LCGFT for Literature Working Group has concluded its work and has been discharged.

- After a discussion at ALA Midwinter on strategies and techniques for retrospectively adding LCGFT headings to bibliographic records that lack them, an ad hoc working
group consisting of Rosemary Groenwald, Yael Mandelstam, and Mary Mastraccio was appointed to deal with the low-hanging fruit by creating a mapping of LCSH form subdivisions and their LCGFT equivalents, if any. Casey Mullin assisted by providing most of the music-specific content. The group also looked at fixed field coding and provided mapping from codes to LCGFT. These mappings are publicly available in the subcommittee’s ALA Connect space:
  
  - 185-155 mapping: [http://connect.ala.org/node/254640](http://connect.ala.org/node/254640)
  - Fixed field mapping: [http://connect.ala.org/node/254639](http://connect.ala.org/node/254639)

Rosemary and Mary will shared the spreadsheet that their group created during a presentation at the Authority Control Interest Group at ALA Annual, entitled “Incorporating the Library of Congress Literature, General and Music LCGFT Terms into the Authorities Database of Your Local ILS.”

**Future of Subcommittee**

After evaluating its charge and reviewing work done up to this point, it was determined that SGFI will continue for likely another year, primarily to work on two projects, to be carried out by working groups:

1) **Working Group on Full Implementation of Library of Congress Faceted Vocabularies.** This group will draft a white paper, whose intended audience will include OCLC, the Library of Congress, the Program for Cooperative Cataloging, and other constituencies, including ILS and authority control vendors, advocating for the full implementation (current and retrospective) of new and emerging faceted vocabularies, most notably the Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT), the Library of Congress Medium of Performance Thesaurus for Music (LCMPT), and the Library of Congress Demographic Group Terms (LCDGT). Casey Mullin will chair this working group.

2) **Working Group on LCGFT for Video Games.** This group will develop a proposed list of video game genre/form terms for inclusion in LCGFT. Rosemary Groenwald will chair this working group. The group will:

   - Identify suitable reference sources for video game genre/form terminology
   - Based on literary warrant as found on video games themselves and in reference sources, and in accordance with principles and policies in the Library of Congress Genre/Form Terms Manual ([https://www.loc.gov/aba/publications/FreeLCGFT/freelcgft.html](https://www.loc.gov/aba/publications/FreeLCGFT/freelcgft.html)), draft authority records for each recommended term, including:
     - Preferred and variant (“used for”) forms
     - Broader and/or related terms
     - Scope notes as needed
Citations of sources consulted
Draft a new section on video games for inclusion in the Library of Congress Genre/Form Terms Manual

Adam Schiff will step down as chair of SGFI and it is likely that Mullin and Groenwald will serve as co-chairs for the coming year.

---

**Linked Library Data Interest Group**

“*OpenVIVO: A Hosted Platform for Representing Scholarly Work*” *(Michael Conlon, PhD, VIVO Project Director, Emeritus Faculty, University of Florida)*

OpenVIVO is a hosted, VIVO system that anyone with an ORCiD identifier can use. Using ORCiD identifiers for sign-on and contributor identification, OpenVIVO can gather works from Figshare, ORCiD, PubMed, and CrossRef. A signed on user can add a paper, or other identified work, to their profile by providing the DOI, along with the contribution they made to the work.

OpenVIVO loads the metadata for the publication from CrossRef in real-time. GRID data is used to identify organizations. An extensive list of journals is included. Data is published to GitHub on a daily basis for anyone to use. Features developed for OpenVIVO will become part of VIVO in future releases. OpenVIVO demonstrates the value of augmentation of the scholarly record with identifiers, the addition and tracking of contribution types, the value of open, immediate reuse of the data through daily export under FAIR (Findable, Accessible, Interoperable, and Reusable) data principles.

OpenVIVO is currently in use at over 140 institutions in 28 countries. Conlon encouraged the audience to join OpenVIVO and experiment with its features. As of June 2016 there were 315 people, over 58,000 organizations, over 44,000 journals, 1270 works, 1717 attributions and 313 research areas. Its data is published hourly at [http://openvivo.org/data](http://openvivo.org/data).

“*Linked Data for Production: Research Questions and Project Goals*” *(Jason Kovari, Head of Metadata Services, Cornell University; Nancy Lorimer, Head of Metadata Department, Stanford University)*

Following the completion of the Andrew W. Mellon Foundation funded Linked Data for Libraries (LD4L) phase 1 (2014-2016), the libraries of Columbia, Cornell, Harvard, Princeton and Stanford Universities along with the Library of Congress partnered on Linked Data for Production (LD4P), a research project investigating linked data in a technical services environment. This Mellon funded effort includes cataloging natively in RDF, data conversion
and developing ontology extensions for the description of art, cartographic materials, performed music and rare materials.

LD4P is focused on immediate needs of metadata production, and is divided into two parts: a community-based collaborative framework, which includes workflows, tools and system configuration that can be shared across institutions; and, institution-based projects, which includes ontologies for specific subject domains, ontology extensions, and workflow development for basic technical services processes. Beyond these two components, there is also the question of non-descriptive metadata and how it will function in a LOD environment and interact with descriptive metadata; these include acquisitions, payment, holdings, circulation, and access data.

The following research areas were briefly described:
- Ontology alignment, including BIBFRAME 2.0
- Ontology extensions for art, cartographic, moving image, performed music, and rare materials
- Tooling and infrastructure (RDF editor, MARC-to-RDF converter (LC tool with pre and post processing))
- ILS functionality; that is to say, the goal is to not have parallel databases
- Link persistence
- Data sharing, or, cooperative cataloging; is there data that is not desired to be shared?

“Linking People: Developing Collaborative Regional Vocabularies” (Jeremy Myntti, Head of Digital Library Services, University of Utah; Anna Neatour, Metadata Librarian, University of Utah)

The presenters began by polling the audience regarding local systems and specifications used for digital asset management and description. This project is warranted on the following premises: data reconciliation among aggregated collections is very badly needed; vendor-based digital asset management systems (DAMS) don’t facilitate authority control very well; applying NACO standards is not feasible, as not all metadata creators are trained in NACO; and, many names of local interest are not seen to be in scope of the Library of Congress Name Authority File.

A pilot project at Brigham Young University used OpenRefine to reconcile name data with the LCNAF and the Virtual International Authority File (VIAF). As was discovered, much manual review and clean-up is needed. Following on this preliminary work, the University of Utah was awarded an IMLS grant titled "Linking People: Developing Collaborative Regional Vocabularies." This project involves four phases: 1) investigating data models to express local/regional name authority data using linked data standards; 2) evaluation of tools used for creating, maintaining, and making this data available; 3) pilot implementation using the
tools investigated in the second phase; 4) assessment of how this type of authority data can improve digital collection metadata on a local, regional, and national level.

**Faceted Subject Access Interest Group**

Eric Childress (Consulting Project Manager, OCLC) began by reporting on the progress of FAST headings being added retrospectively to WorldCat records. So far, over 90 million records have had this enhancement performed. He also announced a new email discussion list, FACETVOC, hosted by OCLC.

Thurstan Young (Collection Metadata Analyst, Metadata Standards, British Library) reported on the British Library’s proposals regarding the Subject Indexing and Classification standards it applies. As background, he explained that the British Library’s acquisitions are up, but cataloging staffing levels are down. In an effort to streamline processing of new materials, BL is exploring FAST implementation in current cataloging of non-copyright deposit materials and in cataloging of backlog materials. One advantage is that FAST genre/form headings are “fully implemented” (as opposed to competing vocabularies such as LCGFT). The primary downside is the fact that FAST is still technically a research project, and so its sustainability is in question.

To assess the community response to these proposed changes, the British Library conducted a survey of stakeholder libraries. Responses came in from libraries in the US, UK, New Zealand and Europe. While responses were mixed overall, arguments against FAST implementation in current cataloging were plentiful. An implementation decision will be made no sooner than Fall 2016.

Lastly, Netanel Ganin (Metadata Coordinator – Hebrew Specialty, Brown University) led a group discussion entitled “‘Jewish men librarians’ is not in LCSH.” He argued for the possible implementation of Library of Congress Demographic Group Terms (LCDGT) for topical use, as it allows identifying multiple personal identities in a post-coordinated fashion. The primary drawback to this approach is the loss of context in some cases where pre-coordination would have provided it.

**PCC Participants Meeting**

*“Panel Discussion: Linked Open Data and the Descriptive Cataloging of Rare Books”* Led by Amy Brown (University of Texas at Austin)

Amber Billey (Columbia University) began by reporting on the work done by the RBMS Controlled Vocabularies Task Force. The Task Force was formed in June 2015 and was
charged to write a white paper on options for ensuring optimization of controlled vocabularies for rare materials in a Linked Data environment. The components they described as required for this process are: hosting and publishing vocabularies on a new domain, launching a new content management system, creating meaningful Linked Data (i.e., ontology data) during this process and providing multiple points of access for this data (including an API, etc.). The Task Force was able to mount an experimental triple store to meet these requirements. A data dump is available on GitHub and feedback is sought at vocabularies@rbms.info. Further work remains, such as reconciling RBMS genre terms against the Art and Architecture Thesaurus, and most importantly, seeking partnerships to ensure sustainability of the triple store and the Linked Data vocabulary. Diane Hillman suggested the Open Metadata Registry for this, though the PCC is a key stakeholder that could also offer support.

“Towards a Framework for Linked Rare Materials Metadata.” Allison Jai O’Dell (University of Florida) reported on the RBMS Data Element Task Force. They were charged to study the discrepancy between granular content standards (e.g., DCRM) and high-level encoding standards. Their proposed solution is a specialized ontology for rare materials description that will enable the full potential of DCRM and controlled vocabularies. She then gave numerous examples of data elements in DCRM that will benefit from the use of controlled vocabularies instead of free-text description. Highlights included: layout, binding term, type, script, colophon and version.

"Poor Jacques... : Encoding Annotations for Rare Book and Special Collections Materials." Joyce Bell (Princeton University) gave a presentation describing the use of the W3C Web Annotations Model in the Linked Data for Production Project (LD4P) at Princeton. They used Jacques Derrida’s personal library as a testbed for encoding dedications (which in this case were much more extensive than simple “From/To” statements) using the W3C model. In this project, 500 object descriptions will be encoded using BIBFRAME and enhanced using entity recognition and standardized data encoding. Data from DBPedia will be leveraged as well. The data set will be made public.

Submitted by Casey Mullin, Chair, MLA-CMC Vocabularies Subcommittee