Subject Analysis Committee


Vizine-Goetz described a project at OCLC to match FAST headings to Wikipedia articles, and publishing those mappings in MARC and as Linked Data. FAST headings are currently present in over 100 million WorldCat records. Their authority records contain 7xx fields for expressing relationships to equivalent LCSH headings. The current project adds an additional 7xx field (the 750 field, to be exact) to the FAST authority record that includes URLs for the Wikipedia and Wikidata instances of the concept (the latter a Linked Data instantiation of the former).
Wikipedia is apt for FAST mappings, as both Wikipedia and FAST operate under compatible systems. Like FAST (which is based on LCSH), Wikipedia is comprehensive and curated, and concepts are based on literary warrant. Per Wikipedia guidelines, good article titles are recognizable, precise and use natural language.

In the project, 180,000 unsubdivided FAST topical headings were matched against 5 million Wikipedia articles. A match was considered valid if either the article title itself, one of its cross-references, or a redirect to the article matched the FAST heading. Wikipedia “infoboxes” were incorporated into the matching algorithm as well; the entity type included there can be used to disqualify articles for non-topical entities such as airports and airlines. Stemming of terms (i.e., stripping words of suffixes and plural markers) was also utilized to increase the match rate.

Approximately 76,000 FAST headings were matched to Wikipedia articles in this fashion. Some unsuccessful headings included those for compound concepts (e.g., Librarians in literature, Libraries and abused women). Not all matches were correct, and require human review in order to refine the process. For example, the Wikipedia articles for Rapness (a place in Scotland) matched the FAST heading Rappers (performers).

Additional tools are available to improve this matching process. LC Classification data present in FAST records could be used to help evaluate Wikipedia matches. Also, some Wikidata data graphs include links back to FAST, but unfortunately these are not always accurate. Further, it is theoretically possible to match FAST terms to sections of Wikipedia articles in cases where the FAST concept is more granular than the Wikipedia concept. Lastly, the project could be extended to other types of FAST headings, such as chronological (whose concepts are well represented in Wikipedia; for example, each year has its own article).

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

Janis Young was unable to attend the Midwinter meeting. Her report was shared with SAC in advance, and passages of interest are excerpted and given below.

Headings for Sexual and Gender Minorities

The hierarchies for approximately 70 classes of persons headings qualified by gay, lesbian, bisexual, transgender, and transsexual have been revised to bring them into agreement with general LCSH practice, and the BTs will now reflect the unqualified heading (e.g., Lesbian authors has the BT Authors).

Formerly, such headings often were assigned numerous BTs that created deep and sometimes uneven hierarchies. A heading qualified by lesbian, for example, may have been assigned BTs that were qualified by lesbian or women or gay, or all of the above.
Classification of Social Media

In January 2018, a new classification range, PN4550-4583 (Social media), was approved. Works about the content of social media and collections of excerpts from social media, including the content of individual social media platforms, will be classified there. Works about the technological and sociological aspects of social media will continue to be classed in the T and HM schedules, respectively.

Genre/Form Terms

Artistic and Visual Works

In February 2018 the Policy and Standards Division (PSD) will approve approximately 50 new LC genre/form terms for artistic and visual works. This genre/form project is a collaboration undertaken by PSD and the Cataloging Advisory Committee of the Art Libraries Society of North America (ARLIS/NA). PSD wholeheartedly thanks the members of the Committee for their time and effort. The proposed terms appear on Tentative List 1802a (http://classificationweb.net/tentative-subjects/1802a.html), to be approved on February 16, 2018.

The proposals represent broad categories of artistic and visual works and can be readily used in general libraries that need to provide high-level genre/form access to their collections of visual works. They do not attempt to replicate the breadth and depth of specialized vocabularies such as the Thesaurus for Graphic Materials and the Art & Architecture Thesaurus.

Materials that are acquired for art collections are usually visual in nature, but are not always inherently artistic. That is, the materials are not “art for art’s sake” but were instead created to serve an informational, documentary, or other purpose (e.g., architectural drawings, trading cards, photographs). The practical need for terms that describe materials that are not artistic in the narrow sense led the Cataloging Advisory Committee and PSD to determine that the highest-level broader term for those terms that represent materials that are not inherently artistic should be Visual works.

This decision required the reconsideration some of the existing hierarchies in LCGFT, because numerous approved terms describe visual works. The current top terms Motion pictures, Television programs, and Video recordings will become narrower terms under Visual works, as will the high-level term Maps. The hierarchies for several other individual terms that refer to visual materials will also be adjusted (e.g., Worm’s-eye views). Those proposed revisions appear on Tentative List 1802a along with the proposals for the new terms.

Demographic Group Terms

Phase 3 of Development
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.

Phase 3 of the vocabulary’s development, which began in January 2016, was intended to test the policies governing the structure of LCDGT. Proposals were therefore accepted for new and revised terms, inasmuch as they were needed for new cataloging. Sufficient proposals have been received to allow a thorough evaluation of LCDGT’s structure and principles, so Phase 3 has ended as of January 2018. Additional proposals will not be accepted until further notice.

LCSH Treatment of Illegal Aliens (discussion at SAC meeting)

The effort to replace the pejorative LCSH heading Illegal aliens has reached an impasse, with no action taken at LC since the subject heading change proposals were publicized in early 2017. SAC members discussed that grievances about the delay should not continue to be directed at LC liaison Janis Young, who has limited authority to enact the change. Rather, continued lobbying for appropriate terminology changes should happen at a higher levels within ALA/ALCTS and LC, perhaps even involving ALA’s legislative liaisons, since the logjam seems to originate from within Congress itself. Daniel Joudrey (Simmons College) will draft a white paper addressing this impasse, which SAC will approve. A possible alternative way forward is for cataloging agencies to apply non-LCSH terminology to works about undocumented immigrants and similar concepts. Such terminology could be sanctioned by PCC, SAC or another body.

Report from the IFLA Liaison

Report from 2017 Session in Poland

The 2017 World Library and Information Congress (WLIC) was held in Wroclaw, Poland, from August 19-25. The Standing Committee of the Subject Analysis and Access Section (SA&A) held two open sessions on Monday, August 21. The first program (9:30-11:30 AM), co-sponsored and organized with the Bibliography Section, was entitled “Challenging Society and Naming Identity: Subject Access and Bibliography in a Multicultural World.” Moderated by Standing Committee member John DeSantis, the program included the following four presentations:

“Knowing Nearby Others: Approaches to Precise Terminology in New Slavic Vocabularies”
by Jean Dickinson.

“Sami Bibliography” by Kari Mathisen and Siri Gaski.

“Genre Theory Applied: Genre and Form Terms in the National Library of Poland Catalogue”
by Kamil Pawlicki.
“Retrieving Oregon Indians from Obscurity: A Project to Enhance Access to Resources on Tribal History and Culture” by Richard Sapon-White.

**Upcoming Programs in 2018**

The next IFLA Conference will be held in Kuala Lumpur, Malaysia, from August 24-30, 2018. SA&A will be holding both an open session program and a satellite conference during the Congress. The title of the two hour open session is: “Transforming Libraries via Automatic Indexing: The Impact on Metadata Creation, Discovery, and Staffing Decisions”. Athena Salaba is chairing the program committee for the open session. The deadline for submitting a detailed abstract is February 20. The full call for papers is available at: [https://2018.ifla.org/cfp-calls/subject-analysis-and-access-section](https://2018.ifla.org/cfp-calls/subject-analysis-and-access-section).

Immediately prior to the conference on August 22, 2018, there will be a full day satellite meeting entitled “Metadata Specialists in the Machine Age.” This meeting is being jointly sponsored by three IFLA standing committees: SA&A, the Cataloguing Section, and the Information Technology Section. More information on the satellite meeting will be available at: [https://2018.ifla.org](https://2018.ifla.org).

**Report from the Art Libraries Society of North America (ARLIS/NA) Liaison**

The art and visual works terms for the LC Genre/Form Terms have been submitted to LC, loaded into the proposal system, and are now out for comment as Tentative List 1802a. The comment period ends on February 2, 2018. This date should allow for LC to analyze the comments before the ARLIS/NA conference in late February in New York City. The announcement about list 1802a can be found at: [https://www.loc.gov/catdir/cpso/artvisualgenreform.html](https://www.loc.gov/catdir/cpso/artvisualgenreform.html).

The proposal for the top term to be Visual works involves adding that as a Broader Term to a range of terms such as Motion pictures, Video recordings, and Maps. Art is also a Narrower Term under Visual works.

In the near term, we did not propose Reproductions as a term. We will continue to look at this issue as we gain experience with using LCGFT for art. We are looking forward to the editorial discussion of the terms. A recent Cataloging Advisory Committee email exchange has already discovered a term that we discussed but abandoned as inappropriate for LCGFT. Now, one of the CAC members has cataloged an installation and could use the genre/form term Installations (Art). We await the hoped-for acceptance of the records and the opening of LCGFT to proposals for additional art terms. During the years we've been working on the terms, we did add a few which came to SACO and were referred to us for possible inclusion in the art project.

The Cataloging Advisory Committee has been working with librarians at Columbia, LC, and elsewhere on the ARTFRAME project which is part of LD4P. Some of us will be attending a
workshop at Columbia in mid-January. The "divided world" (SAF/NAF, SAC/CC:DA) plays out in different ways as we look at how we catalog library (LAM) resources. ARTFRAME includes subject cataloging in its purview but most of the discussions are more descriptive.

SAC Subcommittee on Faceted Vocabularies

The new SAC Subcommittee on Faceted Vocabularies (SSFV) held its first face-to-face meeting at ALA Midwinter in Denver. Several members, including chair Lia Contursi, were not able to attend, so Casey Mullin led the meeting. The group’s discussions focused on the following topics:

Brief progress reports on faceted vocabulary application and development in specialist communities

Music

Mullin gave a brief report on recent MLA efforts, including the recent revisions to the LCMPT and LCGFT best practices documents, and the project to develop and test the algorithm for deriving faceted data from LCSH headings for music.

Video Games

Rosemary Groenwald gave an update on the video games genre/form terms project. The following is excerpted from the working group’s proposal to OLAC:
“Over the course of the last 18 months, the working group has
1) Compiled a list of potential video game terms based on popular usage and/or usage in the industry
2) Researched the potential terms and narrowed the list down
3) Identified suitable reference sources for the video game genre terminology
4) Drafted complete authority records for approximately 80 terms based on literary warrant as found on video games themselves and in reference sources
5) Held ongoing discussions regarding issues/concerns/problems with the treatment of many of the specific terms when creating authority records for them.

The goal is to finish up the creation of all of the authority records by ALA Midwinter 2018. If not by then, the goal is to complete them shortly thereafter.

Although PSD did say that “we (PSD) will be happy to use the working group’s vocabulary as a primary reference source in a future PSD project to add video game terms to LCGFT”, Janis is unable to hazard a guess as to whether or not it would be feasible for LC and PSD to take on an lcgft video game project anytime in 2018. Since it doesn’t appear that PSD will be able to work on this project in the foreseeable future, it makes more sense to adopt an alternative approach like we mentioned in the original charge of the working
group. Especially since much of the work to create a video game vocabulary has already been completed by the working group.

Proposal

1) OLAC takes “ownership” of the video game vocabulary created by the CAMMS/SAC/GFIS/Video game genre working group.
2) OLAC/CAPC requests The Library of Congress to create an "olac" source code to be used with the OLAC video game genre terms when they are used in bibliographic records in the 655 field.
   Example: 655 7 $a Sports video games $2 olac
3) Make the authority records for the video game genre terms available for download from the OLAC website”

Vendor Community

MARCIVE has copies of all vocabulary authority files in their internal database, and is incorporating those vocabularies into regular client project work, which includes enhancing legacy data. Clients have a need to be reasonable in their expectations, considering not all legacy data is equally amenable to retrospective enhancement with faceted terms. MARCIVE has been developing their own LCMPT generation algorithm, and plans to test MLA’s algorithm and offer feedback after MLA releases it.

Backstage Library Works has been adding LCMPT and LCGFT terms for clients, and has undertaken a pilot project to generate LCGFT terms. Messy legacy data poses a major limitation to these efforts.

CJK (Chinese, Japanese, Korean language resources) Community

SSFV members Charlene Chou and Lia Contursi will lead a group discussion on CJK terminology at the Council on East Asian Libraries (CEAL) Cataloging Workshop on March 20, 2018, and will give SSFV a status report afterwards.

Art

As a guest in attendance the SSFV meeting, Sherman Clarke of the Art Libraries Society of North America (ARLIS/NA) referenced his report to SAC (see earlier section of this report for details).

Literature

As a guest in attendance the SSFV meeting, Adam Schiff at University of Washington is working on manual retrospective implementation of faceted terms as time permits (e.g., adding LCDGT terms like Nobel prize winners in 386 fields; adding newly-approved LCGFT terms to older records, etc.)
General Terms

The previous subcommittee (SGFI) prepared a mapping of LCSH subdivisions to LCGFT terms. It was inquired whether this mapping was ever released publicly, and whether SSFV should do so, updating the mapping as needed.


The co-chairs of SAC received responses to the white paper from OCLC, ALA’s Committee on Cataloging: Description and Access (CC:DA), the National Library of Medicine (NLM) and the Program for Cooperative Cataloging (PCC), and shared them with SSFV. SSFV discussed these responses via email prior to the Denver meeting, and continued discussions at the meeting.

Some of the responses seem too have been written quickly, focusing on negative aspects, including limitations and drawbacks of full implementation. Both CC:DA and OCLC mentioned the need to be aware of non-English-language faceted vocabularies; how would SSFV respond to that? In addition to full implementation of LCGFT/LCDGT, should the scope of SSFV be expanded to include these non-English vocabularies or multilingual thesauri such as Getty’s Art and Architecture Thesaurus (AAT)? Should SSFV review the IFLA survey on Genre Form Practices in National Libraries as a reference document?

The group discussed whether and how to respond to these official responses. PCC’s response was brief, owing to the strategic plan that was in the drafting stages at the time the solicitation for feedback was sent by SAC. Since their initial response was issued, this draft strategic plan has become available for comment. Happily, one of the strategic directions PCC has articulated (“SD5. Expand use of linked data value vocabularies to augment, and where feasible to replace, existing metadata practices”) cites the SAC white paper as a useful starting point. This is perhaps the most encouraging (if indirect) feedback SAC has received.

Should SSFV wait for a second, direct PCC response, as they promised? Or can the draft strategic plan be construed as that second response? SSFV members expressed that a response to PCC, in light of SD5, is warranted. Could SSFV be identified as a collaborator in this area? In particular, response could mention a wish to proceed with best practices for NACO records, with their collaboration/coordination.

Beyond PCC, SSFV needs to acknowledge cataloging agencies that lie outside of the purview of PCC. What is OCLC’s role in disseminating knowledge to that broader constituency? And what about cataloging/metadata agencies and communities that might not be inclined to implement LC faceted vocabularies at all?
Who exactly is SSFV’s intended audience, broadly speaking? Three possible groups are: (1) likely implementers of LC vocabularies; (2) people “on the fence” who seek more information; and (3) unlikely implementers (e.g., NLM) who should still be aware of vocabulary application practices they are likely to encounter in metadata in shared environments.

Are there general best practices that are applicable all faceted vocabularies (e.g., no subdivisions are allowed)? Should SSFV produce content that is this general?

CC:DA’s response was thorough and well thought-out (despite the quick turnaround time), but SSFV members did not find it universally constructive to the group’s efforts. Perhaps SSFV’s response to them can be brief, primarily expressing gratitude for the feedback, since SSFV feels there is no need to “defend” their positions to CC:DA. It might also be prudent to acknowledge to CC:DA that SSFV is SAC is not totally ignorant about the areas of concern they outlined. There is some overlap between the scopes of SAC and CC:DA work, but on the whole the two groups work in parallel environments.

An SSFV member who could not attend the meeting and submitted her comments, Charlene Chou, expressed that since these external groups shared their concerns with SAC, SSFV should seek partnerships with them to do pilot tests (in the case of NLM), usability studies and focus group studies (in the case of OCLC), etc. In her opinion, it’s important to get more buy-ins through convincing and powerful use data to support our recommendation for full implementation. In a word, let data speak. Furthermore, OCLC would be an important partner for full implementation. When OCLC was asked if they plan to control other faceted vocabularies (e.g. Getty’s AAT) in Connexion during the Virtual AskQC office hours on January 31, they said it’s on their short list and may talk to Getty in the near future.

The role of the Library of Congress Policy and Standards Division (LC PSD) in SSFV’s work on training and best practices

The group discussed their desired relationship with LC PSD going forward. Currently, the SSFV roster includes no LC staff members (compare SGFI, which included Janis Young from PSD). On the one hand, best practices and other work products will be approved via SAC, where the LC liaison relationship exists; Janis Young or others from LC can comment on SSFV work through that channel. On the other hand, SSFV will greatly benefit from having an LC representative within their midst, so the group can keep tabs on LC’s own implementation progress and plans. The group feels the SSFV chair should approach LC PSD about appointing a representative.

What happens if SSFV wants to promulgate practices that contravene LC practice? For example, bi-level assignment of demonym terms for creator/contributor or audience characteristics (at both the state and national level); LC suggests only most specific term. The recently-announced moratorium on LCDGT proposals also has a bearing on SSFV’s work in this area (see LC Liaison report earlier in this document).
Work plan for best practices and training efforts

The group agreed that a roadmap for best practices work is necessary before substantive work begins. SSFV could start with general best practices (LCGFT general terms, for example), then drill down into specific areas. As needed, sub-groups could be formed, drawing from individuals within SSFV and external to the group if special expertise is needed. SAC will be the venue for soliciting volunteers for sub-groups.

Demographic terms offer the most expansive opportunities, but need more thought as to what’s possible and practical in a production environment (“essential/core elements”).

Casey and Rebecca Belford (his successor as MLA Liaison to SAC) will continue to keep SSFV advised regarding work in the music domain. Lia Contursi and George Prager will keep SSFV advised on efforts within the American Association of Law Libraries (AALL), including any best practices content. SSFV’s role in creating best practices content for law resources is not yet clear.

SSFV will need to engage a member of the cartographic cataloging community, to ensure their needs are taken into account vis a vis faceted data.

Charlene Chou can work with CEAL on CJK issues. She also had this to say regarding art terms:

Should catalogers propose a new art term for LCGFT if a term exists in AAT already? According to the LC memo, “LC proposals represent broad categories of artistic and visual works and can be readily used in general libraries that need to provide high-level genre/form access to their collections of visual works. They do not attempt to replicate the breadth and depth of specialized vocabularies such as the Thesaurus for Graphic Materials and the Art & Architecture Thesaurus.” The definitions of “broad” categories and “general” libraries are quite vague and impractical for catalogers. I will propose using an existing AAT term rather than proposing a new LCGFT since AAT has clearly defined structures with multilingual terms/interface/linked data and is widely used in the international community already.

Regarding the use of non-LC vocabularies: SSFV may want to address when to encourage SACO proposals, vs. when to use an alternative (often more specific) vocabulary. It can be difficult to determine how specific is too “specific” for LCGFT. In general, it is important for SSFV to acknowledge value of external vocabularies.

Scope of best practices work

SSFV will need to determine the scope of its work. Its best practices materials could serve to extend LC manuals, including the Descriptive Cataloging Manual section on authority records (DCM Z1). In doing so, SSFV can also “boost the signal” of LC’s manuals, particularly
since they are new. It will need to be determined which content should be issued as “supplements” to these manuals, and which content should be framed as revision proposals to the manuals themselves.

Lastly, the group touched on other facets beyond genre/form, medium of performance and demographic group terms. These include chronological and geographical data. Best practices content will need to be developed for the use of MARC fields 046, 257, 370, 377, 388 and 648, in MARC bibliographic and authority records as applicable. The roadmap for SSFV’s work should include these components.

Faceted Subject Access Interest Group

Kelley McGrath (University of Oregon). “Using MARC facets with Primo: Strategies and Challenges”

McGrath demonstrated the potential for user discovery of music resources through faceted access. The new LC faceted vocabularies (including LCMPT, LCGFT and LCDGT), combined with dedicated MARC fields for carrying this and other faceted data (including geographic and chronological), has enabled an increase in recall and precision for retrieval of music resources. McGrath began by evoking the anecdote of the frustrated saxophone professor who must navigate the rigid left-anchored structure of LCSH music headings in order to get a comprehensive view of what music the library has for her instrument. In this legacy environment, not even a single keyword will retrieve all relevant hits, since relevant LCSH headings can have the term “saxophone” either expressed in the singular or plural form. The 048 field has been used in some cases to encode and index medium of performance data, but its use among catalogers is not consistent, and the MARC codes given there are not sufficiently granular to bring out specific types of saxophone, for example.

LCMPT and the 382 field offer new possibilities for retrieval of music resources by medium of performance, but it currently suffers from the same “chicken and egg” problem that the 048 field has: system administrators are reluctant to implement the field without its presence in a preponderance of records in the database, and some catalogers are disinclined to spend the time and effort inputting it if their local system is not using the data for retrieval. Retrospective algorithms to enhance older metadata will eventually address this problem, and McGrath has proceeded with enhancements to the Orbis Cascade Alliance’s shared Primo test discovery environment (its “sandbox”).

McGrath and colleagues have developed specifications (or “normalization rules”) for a new suite of facets that expose work/expression data in MARC fields 046, 382, 383, 384, 385 and 386 fields, which complement the existing facet for genre/form (which includes data from MARC field 655). The data in 382 fields accommodate three separate facets, in fact: “Instrumentation includes” (listing individual medium terms in isolation, designating soloists as such), “Music: Number of Performers” (for those 382 fields where $s$ is given), and
“Instrumentation Statement” (showing the entire contents of 382 fields). Additionally, they developed facets that include personal names coded as performers and composers, employing advanced MARC logic to achieve this limited scope.

While these new facets serve as proof of concept of the possibilities of comprehensive faceted access to music resources (and by extension, all library resources), several present-day limitations pose barriers to success. First, since not all bibliographic records yet carry these faceted data fields, recall will be incomplete, thwarting user expectations. Second, LCGFT and LCMPT are built with strict hierarchies, but Primo does not yet offer the user an option to search on a broad term (e.g., Popular music) and retrieve records that carry terms that are narrower to that broad term in the hierarchy (e.g., Rock music). Third, authority data for musical works that carries faceted terms for attributes of those works could offer an alternative method for retrieval of those works (obviating the need to encode the same work data in each bibliographic record); alas, Primo and other popular discovery products do not currently offer the ability to incorporate authority data in retrieval in this way.

Fourth, compilations of works with disparate attributes lead to potential false hits, since faceted terms are not currently linked to the work they describe. For example, a sound recording including a concerto for piano and a sonata for violin will be retrieved if the user searches for Concertos + Violin. Fifth, discovery implications of demographic characteristics have yet to be studied extensively. Should creator characteristics be encoded only once, in the authority record for the person? One argument against this approach is the case of the child composer. How should this attribute be brought out? At the bibliographic level for works composed during their childhood? At the authority level for those same works? This cannot be done in the authority record for the person, since composers that survive into adulthood cannot be said to be “children” throughout their lifetime. Lastly, the medium of performance statement facet is also subject to recall problems, since medium components may be input in differing orders in different records. For example, a 382 field for a piano quintet may give piano first, or last.

Despite the aforementioned difficulties, McGrath encouraged the audience to explore the discovery possibilities offered by these new facets, and invited people to visit the OCA’s Primo sandbox (http://alliance-primo-sb.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vid=WWU) and submit feedback.

Diane Vizine-Goetz (OCLC Research). “Update on FAST”

Diane Vizine-Goetz presented briefly on a new feature in the FAST suite of tools called importFAST. It is an experimental tool that allows users to propose new FAST headings, including name headings adopted from NAF. OCLC is reluctant to add the entirety of NAF upfront, since there are many names that will never be used as subjects. Proposed FAST headings must be based on LCSH headings. They can be simple headings or headings with subdivisions. The proposer can include notes to alert OCLC to unusual situations.
OCLC is also reformatting about 5,400 non-conference event headings, changing the MARC tag from 111 to 147. Lastly, some FAST form headings have been deprecated in favor of LCGFT terms.

For a description of the Wikipedia/Wikidata FAST project, see the earlier section in this report.

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