To: Jennifer Bromley
        Rocki Strader
        Co-chairs, ALCTS CaMMS Subject Analysis Committee

From: Tina Shrader, Chair
        ALA/ALCTS/CaMMS/Committee on Cataloging: Description and Access (CC:DA)

Subject: CC:DA Feedback on the White Paper "A Brave New (Faceted) World: Towards Full Implementation of Library of Congress Faceted Vocabularies"

CC:DA found the white paper helpful and informative and is glad to see the focus on improving access to non-topical aspects of library resources. We comment below on the main sections of the document followed by our responses to the questions from SAC.

Training:

In order to be optimally useful, these facets not only have to be populated as comprehensively as possible, but also as consistently as possible. That requires wide dissemination of knowledge about how to apply these vocabularies.

CC:DA concurs that training is key to such a large and significant transition. This training needs to be backed up by effective documentation and policy, as well as community buy-in to a common approach. Although documentation is never finished and new needs may emerge over time, documentation that is comprehensive and stable enough for implementation is necessary before embarking on training.

In addition to technical matters of application and consistency, we need to highlight why change is needed. Training should begin at a broad level and emphasize why we need to recondition our data so that it can reach a wider audience and take better advantage of current and emerging technological capabilities, such as linked data. Additionally, some level of training should be provided to non-cataloging staff (such as reference and public services staff) who will be working with the faceted data at the front-end.

Use of faceted vocabularies in work and expression authority records:

There are difficult decisions to be made about where these vocabularies should be used: in bibliographic records, in authority records or in both? The rationale behind the recommendation to create work and expression authority records only for works that appear or are likely to appear in multiple manifestations is unclear to us. In this age of online versions and technological change, how many commercially-distributed resources are not likely to end up embodied in

multiple manifestations? It doesn't seem like it would often be safe to make the assumption that something will be issued in only one manifestation even if there is only ever one expression.

Some of us argued that a more useful goal might be to strive for some sort of minimal work and expression stub for every manifestation. Recording this kind of information redundantly in bibs at the manifestation level has obvious drawbacks: not only extra work, but variation in completeness and accuracy.

Unfortunately, most current systems cannot make good use of data in authority records. In addition, many institutions and catalogers do not have the means, permissions, or training to create shared authority records, and those who do may be hard put to create authority records for everything they catalog. This approach would also mean yet another layer of training, since most catalogers do not deal with title access points. There doesn't seem to be a practical path to creating authority records for everything or even most works and expressions. Putting information from these new vocabularies in bib records gives us a place to record this information now. That data can later be converted to another form, such as linked data, which doesn't require authority records as we know them, particularly for works and expressions. However, even if we focus on using these new terms in MARC bib records rather than authorities, those with the ability to create authority record should be encouraged to add them to authority records in situations where the record otherwise can only be disambiguated using information in 670.

Recording data in either bib or authority records has clear drawbacks. We need practical ways to more widely share the creation and maintenance of work and expression data, expanding the range of catalogers who can contribute while also safeguarding quality control. Perhaps there is a third way for the short-term: creating some sort of external database of works and expressions that is more open to editing than the NAF and that can be linked to bib records or clusters of bib records. These external records could initially be derived from information in bibs. They could then be used to programmatically populate and update bibliographic records since the values from these vocabularies are stored in discrete MARC fields. In this scenario, although the data would be stored in bibs and carried with bib records, it would be maintained more parsimoniously in a single location. If we can manage data at a more granular level, perhaps it would make more sense to allow provisional assertions from a broader community (or algorithmic generation) with a more limited pool of people able to vet and somehow approve statements.

Many of us also work extensively with digital repositories and digital asset management tools. Most of these systems are not designed to support authority control and/or identity management principles. These new faceted vocabularies could also be useful in this environment. It would be helpful if there were tools and guidance that would support the use of faceted vocabularies in these non-MARC-based systems.

Retrospective conversion:

Clearly, much retrospective population of these new fields can (and should) be done by automated means. However, some task group members expressed concerns about the accuracy of the underlying data and the potential for incorrect inferences. There are many miscoded records, variations in interpretation and practice, and records with incomplete information. Records may be ambiguous or contradict themselves; some situations cannot be resolved without consulting the resource and, even then, there will be cases that aren't clear-cut. It would be ideal if there were some way to mark that a human had reviewed the assignment of these new terms, either with or without the piece in hand. Unfortunately, in the MARC environment, this is only practical in authority records.

Analogous to the way that law libraries have formed a working group to create a cumulative list of MARC records to add specific legal genre/form terms to, a group of libraries (e.g., public libraries) might be able to come up with a list of records for at least the largest literature genres. These genres are often shelved separately and genre information could be extracted from item records.

In the discussion of mapping 008/33 to genre terms (p. 25), there is a recommendation that the general term "Letters" be established. Is there a reason the recommendation is for that term rather than "Correspondence"? Correspondence would fit more appropriately in the hierarchy of existing narrower terms, such as "Personal correspondence" and "Business correspondence." It's also a term commonly used in digital repositories and other environments to describe letters.

The example given on page 21 for changing $v Juvenile films into Children as audience and Motion pictures as genre/form is in conflict with the scope note for "Motion pictures" that says to use for "Collections of films that are composed of multiple genres and/or forms to which more specific headings such as Nonfiction films or Comedy films cannot be applied." Other instances where broad terms are added retrospectively are also likely to be in conflict with what catalogers are told to add manually.

In some ways, it might actually be a good thing to add a subset of broad terms to all records in addition to more specific terms. In a faceted interface, it can be useful to be able to adjust the level of specificity of a query in different situations. Some users may only be interested in very specific types of poetry, such as haiku, but others may be interested in seeing all the 20th century Japanese poetry or all the poetry in a less common language. Although systems should be able to use the hierarchy encoded in authority records to provide this kind of navigation, it is unlikely that many systems will be able to do this in the near future. Assigning a general category term in addition to more specific terms would lead to more consistent results for users in most systems, given that many records will only have broad terms that were assigned retrospectively.

Display and indexing in bibliographic records:

There were some questions about the usefulness of these facets in discovery interfaces where keyword searching is dominant and records come from many sources that do not use consistent controlled vocabulary. It would be useful to look for research or perform user studies to investigate this.

Much of the data in these new fields duplicates data that already appears in other places in the bibliographic record. Particularly if data is retrospectively added to bibs, this may lead to displays that are "cluttered" with redundant information. Many discovery systems do not have tools to perform the sometimes complex formatting, such as adding punctuation, that is necessary for clear display of some of these new fields. Systems often index terms by their strings, not vocabulary (or URI) so that a listing of a term may include several different vocabularies.

ILS's and other cataloging tools do not sufficiently support these fields for efficient input and data quality control. Some ILS's do not recognize the $2, and thus are unable to differentiate between different vocabularies in a 655 _7, without some local hacking to authority files. Most do not have the capability to control multiple occurrences of vocabularies in a field, as in the 382. OCLC's Connexion, which is used by many catalogers, is able to control LCGFT terms in 655. Data from the other vocabularies would be more consistent and accurate in the shared WorldCat database if Connexion were able to control data in those fields as well.

Display and indexing in authority records:

It is generally more logical and efficient to store the creator characteristics with agent entities and the work and expression attribute with those entities rather than associating then directly with manifestations. However, certain types of attributes introduce complexities. For example, some characteristics are time-limited, e.g., only some of Mozart's works should be identified as being created by a child. Some links also appear redundant, such as LC's example of poetry being written by poets. If you don't mark all poetry as being by a poet, where do you draw the line?

General comments:

It might be helpful to include a one-page list of acronyms used in the document.

We were surprised that there is no discussion about current implementations of LCGFT for literary works and general materials, areas which will likely affect the largest number of catalogers. The general historical description of LCGFT seems slight in comparison to the others as well (with no mention of partnerships as with the LCMPT history, though there were several), and the quoted statistics are from over a year before this paper was issued.

Although LCGFT terms are supposed to describe a single concept, there is a sense in which phrases like "Western television programs" combine two things, Westerns and television programs, that might be usefully deconstructed. LCGFT divides moving images based on their initial method of distribution: film or video vs. television. Although this is an important distinction, the "television program" headings give the impression of describing television series and are less intuitive when applied to made-for-TV movies. From the other direction, strictly speaking, series produced for the internet are not television series, but there is a temptation to apply television program headings and these productions can sometimes win awards given for television.

Moving images were the first area where LCFGT were established and it may be a good idea to review them in light of subsequent developments.

Copy Editing Notes:

There is a typo at the end of the first full paragraph on page 9. In both of the last two sentences, Library of Congress Demographic Group Terms is abbreviated "LCGDT" rather than "LCDGT."

Footnote 15 on page 10 refers to an LC web page with "with free cataloging documentation and training resources". However, the link given is https://www.loc.gov/cds/products/catManual.php, which primarily describes documentation available by purchase through Cataloger's Desktop. If there is something specific in this list that is relevant, it should be explicitly identified. Otherwise, this information should be omitted as it is unhelpful.

Should the recommendation on page 26 for a new LCGFT term for Letters be pulled out so it's highlighted like the other recommendations?

Responses to SAC Questions:

What is feasible in the near term and the long term?

Music and audiovisual catalogers already consistently use LCGFT in their cataloging. In the near term, use should be expanded by strongly pushing for the implementation of literary genres in combination with some retrospective conversion. It is easy for most people to see a need there, and would be a good way to start widespread training.

In the near-term, shared documentation, such as the BSR, should be updated to clearly delineate topical subject and genre as two different data elements.

Basic retrospective addition of LCGFT and LCDGT terms to bib records based on MARC fixed fields and $v mappings is very doable in the shorter term. The more complex conversion of

music terms will have programming in the near future. These algorithms will allow both local catalogs and WorldCat to be populated with terms from these vocabularies.

Basic decisions on usage in bib vs authority records is necessary, with the reality in mind that for most it is impractical to do duplicative work. ILS's need to be queried about their ability to handle multiple controlled vocabularies, and pressed to improve where necessary.

In the longer term, LCSH has to be converted sufficiently to allow for dropping of duplicative entry in LCSH. Best practices need to be developed for issues that are not addressed in the basic guidelines as they occur, and continuous reassessment of both guidelines and best practices need to take place as usage matures. Vocabularies need to move from reliance on unique string matching to matching on an identifier for the term (not the identifier for the authority record).

Concern was expressed about the feasibility of achieving a critical mass of records with Creator Category fields, particularly if its application is limited to a narrow set of in specific use cases.

What concerns or issues do you believe need to be addressed first?

Discovery layer and ILS vendors need to start working on displaying and indexing these fields so that they are ready by the time everyone is trained.

Aggregates present a significant challenge that has been insufficiently discussed in this paper. These are common and with current coding in bib records, are likely to provide misleading results, i.e., combining facets that belong to separate component works. There should be a plan for dealing with these before plunging into implementation.

Although the long-term goal is to remove the legacy LCSH headings, if possible, this should not be done until there is a mechanism for retaining the connections these contain. For example, many LCSH music headings combine genre and medium of performance. This link is broken when using the faceted vocabularies, which leads to a loss of information on records for CDs that are compilations.

It is sub-optimal to begin using these new fields without a way to proactively record these connections. Unfortunately, there aren't any easy options in the existing ecosystem for doing this in bib records. MARC includes $8, which can be used for linking fields, but it is not implemented for bibs by any major vendors, including OCLC. $8 also doesn't help with linking to parts of fields. Ideally, you would want the transcribed title and statement of responsibility (probably part of 505) linked to the field for the preferred access point for the work or expression as well as all the relevant facets.

Alternatively, one could use a bound with-type approach with multiple bib records, but here again reliable technological support is lacking and it complicates record sharing. Also most

systems expect the individual bibs to carry all the general manifestation information such as publication info (which would be redundant in this situation) and pagination (which might be too much work for any benefit). Recording information for the components in authority records is a cleaner solution, but this is limited to those with the ability to edit authority records and is useless in systems that do not take advantage of information in authority files.

What elements within these vocabularies could be considered "core" (if any)?

The more consistently these elements are added when they are applicable, the more useful they will be. The members of the task group expressed varying perspectives on what, if anything, should be core, and no clear consensus emerged. One argument in favor of making fields core is that they will not be useful for searching if they are incompletely populated. Other factors considered include the perceived importance of the attribute for searching, the degree to which the element can be retrospectively populated, and the extent to which the element contains data that has historically been recorded in the record in some form.

Some concerns were expressed about making elements core. These include creating extra work by requiring the addition of broad terms that don't add value (such as "Music", "Informational works") if no other term applies. Making things core may also encourage catalogers to look for and supply data that isn't there, e.g., many materials do not have an explicit audience and trying to find something to record is not an effective use of time. And there will inevitably be resources that are inherently difficult or impossible to fit into the current scheme or for which the effort to determine the correct value is out of proportion to the benefit.

It was also pointed out that it may be desirable to have different decisions about whether an element should be core for bib and authority records.

What policy issues (locally, nationally, internationally) need to be considered?

● The amount of guidance we can expect to receive from LC and whether or not LC will implement use of the vocabularies in their cataloging practices.
● LC, PCC instructions (current and forthcoming) and how they relate to domain best practices in groups already using terms).
● Genre/Form vs LCSH in PCC when they duplicate one another.
● Address other areas of cataloging in which these terms overlap, and whether the duplication is necessary or could be removed.

What suggestions do you have, or actions would you consider taking, for expanding awareness and use of these vocabularies at a global level?

Training needs to be developed and publicized. The PCC Standing Committee on Training would be a good group to take this on. Announcements about the vocabularies and training resources should be posted widely on cataloging email lists and in other relevant forums. These should note which resources are free of cost. Presentations and pre-conferences at state and national conferences, as well as webinars, are good ways to disseminate training.

On a global level, there should be further outreach to non-ALA groups, including those in other English-speaking countries. For non-English-speaking countries, support work to translate these vocabularies or map these vocabularies to others. Such a project has already started for LCMPT.

We have to look very closely at the current policy that terms be recorded in the new vocabularies, but still also in LCSH. This is added work for catalogers, and not negligible over a week's work. Until that perceived necessity is dropped, there will be resistance.

ILS and discovery interface vendors need to raise the sophistication levels of their treatments of multiple controlled vocabularies.

Getting buy-in at the top will make consistent implementation and application more likely. It may be helpful to reach out to library administrators -- particularly those at large, influential institutions -- and explain why faceted vocabularies will lead to better access and retrieval and better use of library resources by both library and non-library communities. Targeting administrators who are professionally involved in ALCTS would be another good option.

What training issues would you be able to address?

The charge of CC:DA does not include the delivery of training, so this body is not equipped to handle training issues at this time. As the use of these vocabularies expands, if there are training-related needs that would be appropriate for CC:DA to address, the committee will be more than happy to help.

What outcomes or deliverables are most desirable?

- Systems should be enhanced so that these faceted vocabularies can actually be useful in the search and discovery environment.
- The big players (OCLC, LC, other national libraries and cataloging groups) need to facilitate the use of these vocabularies through documentation and training.
- Records in bibliographic utilities and LC's catalog should be retrospectively populated with terms from these vocabularies where possible.

- Software or data sources should be available to enable libraries to add these terms to their existing local records.

May we count on your endorsement and participation/collaboration in the implementation of these vocabularies? The Subject Analysis Committee hopes to foster development of best practices for the use of these vocabularies.

CC:DA endorses the implementation of these vocabularies and would be happy to contribute to implementation efforts.