ALAN Annual Conference 2019, Washington, D.C.
Report by Karen Peters (Library of Congress), Chair, Encoding Standards Subcommittee

MARC Advisory Committee Meetings (June 22 and 23)
Over the course of two meetings, the MARC Advisory Committee (MAC) considered three Proposals—all derived from Discussion Papers presented at the Midwinter Meeting last January—and two Discussion Papers. All three of the Proposals passed. The first Discussion Paper will be fast-tracked, while the second will be coming back as a Proposal at the Midwinter Meeting next January. Besides these five items, there was an issue involving one of Proposals approved at the last Midwinter Meeting to be dealt with.

MARC Proposal No. 2019-01: Designating Open Access and License Information for Remote Online Resources in the MARC 21 Formats.
While MAC approved this Proposal at its 2019 Midwinter Meeting, the Proposal was not fully implemented due to issues involving the 856 (Electronic Location and Access) field. To correct the problem, an amendment to the Proposal defined a new $7 (Access status) in all five MARC 21 Formats, which will be used to indicate whether access to a remote electronic resource is open or restricted. The proposed change was approved and intended for publication as an addendum to MARC Update No. 28, which was issued last May.

Update: in July, after the MAC meetings, additional problems with the Proposal were identified, requiring an additional amendment, which was approved (please see MARC Proposal No. 2019-01 for details). On July 12, an addendum to MARC Update No. 28 was published that incorporated all of the changes to Proposal No. 2019-01, as well as adding a change that resulted from a Fast-Track Proposal that was approved after the MAC meeting in June: MARC Fast-Track Proposal No. 2019-FT01 (Adding a Code for Audio Belt in Field 007/01 of the MARC 21 Bibliographic Format). For the addendum announcement, please see https://www.loc.gov/marc/marc21_update28_addendum.html.

This Proposal, from the British Library, recommended that changes be made to the Holdings Format character position 008/06 (Receipt or acquisition status) in order to accommodate online publications made accessible via a third-party platform. The proposal passed, with the incorporation of a change in the wording of the definition for the new code value 6 (External access).

This Proposal, from OLAC, added 2 new subfields to Field 041: $i (Language code of intertitles) and $t (Language code of accompanying transcripts for audiovisual materials). Please note that, by definition, $t is intended to be used for non-musical audiovisual materials only. The proposal generated a great deal of discussion but ultimately passed with one change: the proposed modification of $m (Language code of original accompanying materials other than librettos) to
exclude its use for transcripts was withdrawn, with the result that the original language of transcripts will be coded in $m.

Proposal No. 2019-06: Defining a Field for a Subject Added Entry of Unspecified Entity Type in the MARC 21 Bibliographic Format.
This Proposal, from the German National Library, recommended the addition of a new 670 field (Subject Added Entry—Type of Entity Not Specified) to the MARC 21 Bibliographic Format. After much discussion, it was agreed to designate the new field as 688 instead of 670, since 670 in the MARC 21 Authorities Format has a very familiar but different association (Sources found) for catalogers. The proposal, with the new field designation, was passed with the additional incorporation of a number of minor changes.

This Discussion Paper, also put forward by the German National Library, proposed the addition to Field 751 (Added-Entry—Geographic Name) of a new (and repeatable) $g (Miscellaneous information), a subfield that already exists in the other X51 Fields. It was agreed that the Discussion Paper should be turned into a Fast-Track Proposal; this is likely to be approved soon.

Discussion Paper No. 2019-DP05: Adding Subfield $0 to Fields 310, 321, and 521 in the MARC 21 Bibliographic Format
This Discussion Paper, put forward by NDMSO with the goal of facilitating the conversion of data from MARC to BIBFRAME and back, generated a significant amount of discussion, primarily in regard to the proposed change to the 521 Field (Target Audience Note). While some present felt strongly that the 385 Field (Audience Characteristics), which already permits the use of $0 (Authority record control number or standard number), is the appropriate field for recording this information, others pointed out that some information recorded in the 521 Field, such MPAA ratings, has no place in the 385 Field. After taking the issues raised into consideration, the Discussion Paper’s authors will make revisions and convert 2019-DP05 to a Proposal that will be presented at the MAC Meeting next January.

LC BIBFRAME Update Forum (June 23)
The Forum began with Sally McCallum (Library of Congress) announcing that the Library’s BIBFRAME Works and BIBFRAME Instances files are now publicly available. These can be keyword-searched and are downloadable for systems use (in the formats RDF/XML, N-triples, and JSON) through the Library’s Linked Data Service at id.loc.gov. McCallum urged the audience to make use of this resource, but also cautioned us to remember that it is still a work in progress: while improvements are continuously being made, the data in question are not entirely consistent due to changes in cataloging conventions over the one hundred years or so during which the catalog records providing the data were created, due to issues of data loss suffered when these records were undergoing retrospective conversion, and so forth.

McCallum also announced that the Library’s BIBFRAME Pilot is expanding to include nearly 100 participants, including media catalogers, with training of new participants taking place
beginning in July. Currently, pilot participants are required to catalog materials in both BIBFRAME and in MARC, as the Library will continue to distribute records in MARC even after its transition to BIBFRAME is completed. By this fall, however, it is expected that a BIBFRAME-to-MARC conversion will be implemented, permitting pilot participants to catalog in BIBFRAME only. Note that when the conversion of MARC to BIBFRAME takes place, it will be without the duplicative data found in MARC.

McCallum’s announcements were followed by two reports that focused on the Mellon-Grant funded Linked Data for Production (LD4P) sub-project to develop Sinopia, a BIBFRAME editor for use by LD4P participant libraries that was initially built on the Library of Congress’s BIBFRAME editor. In the first report, Jeremy Nelson and Josh Greben (Stanford University) spoke of their current work with Sinopia, including the goal of incorporate access to external data sources such as Wikidata and ShareVDE, and the expansion of the current cohort of institutions using it. Note Sinopia is being run on Amazon Web Services (AWS). Further information can be found at the Sinopia website: sinopia.io.

The second of these reports, “Questioning Authority (QA) Data Access,” by Lynette Rayle (Cornell University), discussed Cornell’s related activity in the LD4P Project in support of accessing controlled vocabularies from multiple authority sources besides those of the Library of Congress. The speaker invited the audience to try out the resulting Authority Lookup Server at https://lookup.ld4l.org. Slides of her presentation are available at https://tinyurl.com/y3r3vagv.

These two reports were followed by a presentation by Nathan Putnam (OCLC), in which he gave a brief history of OCLC’s activities involving Linked Data and presented its plans for the near future. While OCLC will continue to support other encoding standards, it is actively working on enabling BIBFRAME in WorldCat. OCLC Research is currently experimenting with a BIBFRAME converter and plans to present an update on that project at ALA Midwinter next January. OCLC is also creating a survey on persistent identifiers that they will be distributing in the next month or so. Other projects for the upcoming year include plans to convert Dewey into Linked Data. More information on OCLC research and work with Linked Data can be found at https://www.oclc.org/research/themes/data-science/linkeddata.html.

Finally, Martha Sanders (Innovative Interfaces) gave a presentation on III’s creation of a BIBFRAME-based discovery platform that employs a native Linked Data environment with a visual context that permits retrieval and organization by the user of combinations of subjects, titles, and names. III is working on integrating the platform with additional data sources such as Wikidata, as well as on permitting the creation of “work-level rollups” that would gather together all the different formats of a work.

**OCLC Linked Data Roundtable: Stories from the Front (June 22)**

This edition of OCLC’s Roundtable consisted primarily of reports on three current projects involving Linked Data, in this case with a focus on the use of Wikidata. The first, “Wikidata and Scholarly Communication,” by Anchalee (Joy) Panigabutra-Roberts (University of Tennessee,
Knoxville), involved her research into the possibilities of contributing to and querying Wikidata, which contains links from sources such as ORCID and LCNAF. She began by creating name entities in Wikidata for her own works in order to create an author’s profile as a test case, then used Wikidata tools to query and manipulate the resulting data. She then went on to look at the possibility of using Wikidata to provide better access to the component contents of resources—book chapters, essays in a collection, presentations included in conference proceedings and so forth—as these are not always accessible through bibliographic records. The slides of her presentation are available at: https://zenodo.org/record/3256713.

In the second report, Matt Miller (Library of Congress) discussed Wikidata at the Library of Congress. Miller considers Wikidata, which employs structured data, to be an “onramp” to integrating library resources into the World Wide Web. One way of facilitating this integration is through the addition of identifiers to Wikidata. By reusing existing mappings from OCLC’s VIAF, which includes links to LC Authority records, Miller was able to bulk load approximately 450,000 new mappings into Wikidata and connect these to id.loc.gov, bringing the total of LC identifiers in Wikidata to over one million. He is now beginning to think about how to leverage Wikidata metadata connected to library resources to enhance user discovery and knowledge, and also of adding links to Work as well as to Name authority records. His presentation slides were not available at the time of writing, but his recent blog post, “Integrating Wikidata at the Library of Congress,” deals with the subject of his ALA presentation. See: https://blogs.loc.gov/thesignal/2019/05/integrating-wikidata-at-the-library-of-congress/.

The third report was “Beyond Wikipedia: interconnecting human knowledge through Wikidata,” by Andrew Lih (Wikimedia DC/Metropolitan Museum of Art). Lih noted that, while the goal of Wikipedia is to compile the sum of all human knowledge, the result is limited because its entries are text/language dependent. To get around this limitation, the idea is now to “re-[use], re-[mix], and re-[imagine] human knowledge in new, innovative ways.” This resulted in the creation of new Wikimedia: Wikidata—semantic representations of structured data—and Wikimedia Commons, which includes multimedia content and objects with structured metadata. Lih gave an outline of the progress of these resources over time, pointing out the increase in referenced Wikidata statements and Wikipedia citations in scholarly writing, the rise in the number of user edits to Wikidata per day, and the increase over time in the geographical distribution of the users creating Wikipedia entries and making Wikidata edits. He also pointed out the increase in links to authority sources such as LCNAF and VIAF found at the end of Wikipedia entries (the “Pot of Gold”), the addition of which has served to engage new GLAM (Galleries, Libraries, Archives, and Museums) partners. Lih also discussed the tools—some of which take the form of games—provided by Wikimedia that further permit contributors and the public to help improve its content. Other projects being developed include Crotos, a browser for artworks; WikiCite, an open database of citations; and Scholia, which is similar to Google Scholar, but open. Slides of this presentation are available at: bit.ly/ala19wikidata.

MARC Format Transition Interest Group (LITA/ALCTS) Meeting (June 22)
The Interest Group meeting consisted of two presentations. In the first, “Machining metadata that is good enough: updating The Ohio State University Libraries Discovery platform through
virtual metadata reconciliation,” Terry Reese (Ohio State University) recounted the process begun when he was asked by OSU Libraries to “make things easier to find,” a process complicated somewhat by the Libraries move towards a “mobile first” philosophy. Reese noted that libraries have difficulty with discovery because their resource data is siloed: library data are tied to specific standards that are not always interoperable, and the quality of the data in question can be uneven (he also noted that librarians don’t always think like their users). What was needed is a common data model, and the OSU Libraries are now working towards a platform that aggregates data into a simple data schema that not only supports research, but automatically normalizes and enhances metadata on ingest. According to Reese, it is too early to tell if this attempt will be successful.

In the second presentation, “MarcEdit and OpenRefine for Batch Metadata Transformation,” Brian Clark (University of Illinois at Urbana-Champaign) presented two projects aimed at the exposure of “hidden” microform collections that made good use of these tools. In the first project, the tools were used to create MARC records for a large microform collection of hymn tunes. Data from Nicholas Temperley’s Hymn Tune Index was exported to an Excel spreadsheet, where OpenRefine was used to organize and clean it, and to reconcile names to LCNAF. MarcEdit was then used to create K-level MARC records that were uploaded to WorldCat. The second, similar project discussed involved the creation of MARC records for a collection of Federal Government documents on microfiche.

**ALCTS Metadata Interest Group Meeting (June 23)**

**Metadata Standards Committee (ALCTS-LITA) (June 23)**

The ALCTS MIG meeting agenda can be found here: [https://connect.ala.org/alcts/viewdocument/agenda-for-annual-meeting-1?CommunityKey=45b67b1a-f188-4e33-988f-a491aae3f94c&tab=librarydocuments](https://connect.ala.org/alcts/viewdocument/agenda-for-annual-meeting-1?CommunityKey=45b67b1a-f188-4e33-988f-a491aae3f94c&tab=librarydocuments)

The Metadata Interest Group meeting was split equally between a presentation and a business meeting. The presentation, by Library of Congress’s Rick Fitzgerald and Grace Thomas, outlined the history and present state of the Library of Congress Web Archiving Program, which began in 2000 as a pilot project. While libraries in other countries around the world also do web archiving, they typically have a legal mandate to focus on the web materials of their own country. The Library of Congress, not having such restrictions, collects materials from all over the world, relying on their selecting librarians for appropriate resources and collection building in their specific areas. The speakers outlined the Web Archive’s history, including the technical changes and improvements made to it over time; outlined the current model, which uses MODS; and discussed research projects being undertaken to enhance its usefulness in the future. While the presentation slides were not available at the time of this writing, additional information on the Program can be found at its FAQ page: [https://www.loc.gov/programs/web-archiving/about-this-program/frequently-asked-questions/](https://www.loc.gov/programs/web-archiving/about-this-program/frequently-asked-questions/)

The Metadata Standards Committee meeting was devoted to the continuation of the committee’s ongoing work towards creation of a flexible, schema-agnostic metadata standards
framework. When completed, interested institutions will be able to use the framework to assess metadata standards, both prior to implementation and on an ongoing basis.