**Practical Applications of Linked Data**

**MLA Annual Meeting 2016, Cincinnati, OH**

*Summary written by Keith Knop, Florida State University*

*Session presenters: Kimmy Szeto, Baruch College; Karen Hwang, Pratt Institute/Linked Jazz Project; Steven Folsom, Harvard University (formerly from Cornell University); James Soe Nyun, University of California, San Diego*

The second plenary session of the conference was Practical Applications of Linked Data. Presenters were Kimmy Szeto (Baruch College), Karen Hwang (Pratt Institute/Linked Jazz Project), Steven Folsom (Harvard, formerly Cornell), and James Soe Nyun (University of California, San Diego).

Kimmy Szeto began the session by introducing its overarching theme ( “Are we there yet?”) with an overview of libraries’ current involvement with metadata and the structures we have built to support it. As we include an ever-growing number of digital objects in these structures, designed largely with physical items in mind, the weaknesses of the system have become apparent; in particular it has become obvious that much of our information remains quarantined away from the wider internet. Linked data is a solution, providing web access to data in a structured, standardized way.

BIBFRAME is one library response to the need for linked data, but many of our projects go beyond simply bibliographic information. Accordingly we turn to other schemas to fill the gaps, using the mixing and matching allowed by standardization to custom tailor our metadata to our projects. “Data curation” is in large part an act of adjusting these various schemas so they play well together, as the next two presentations illustrate.

Karen Hwang presented on Linked Jazz (linkedjazz.org), a project of the Pratt Institute School of Information that began in 2011. The jazz world features a complex web of personal and professional relationships documented variously in photos, letters, liner notes, and oral histories, and other resources; the Linked Jazz project is an attempt to document those relationships using linked data, enabling researchers to dynamically explore the social networks of jazz artists.

The current nucleus of the project is a set of over fifty digitized oral history transcripts drawn from multiple universities and institutions. A transcript analyzer program uses natural language tools to match names against a preloaded list of 9,000 musician name authority records and create basic relationships between them and the speaker being recorded. Partial names or names with no matches can then have authority data linked or created by the user.

Linking the names to an authority file allows related data to be brought in from other open data resources, such as DBPedia (a project that extracts structured, machine-actionable data from Wikipedia) and MusicBrainz (an open music metadata project). For example, between DBPedia, MusicBrainz, and the Virtual Internet Authority File (VIAF), the project was able to pull in gender data for over 70% of the names in the Linked Jazz musician list. Each resource can be used as a stepping stone to acquire data from additional resources. Ongoing projects include development of a linked jazz ontology, incorporating links to collections at Tulane and the Library of Congress, and pulling in additional data about musicians such as instruments played.

The Linked Jazz dataset, along with an API and query information, is available for researchers and developers to download and explore, but users can also interact more casually with the data using the network visualization tool (linkedjazz.org/network), which shows the relationships between musicians as a graph. Hovering over a musician’s portrait gives access to a Wikipedia link and a sample of their music, where available. Clicking on a portrait displays only the relationship graph for that musician, with links to the transcript sections that detail the relationship and links back to the complete transcript. Also available to the general public is the 52nd Street crowdsourcing tool, which allows users to read transcript sections and contribute more specific relationship data than the default “knows of” relationship.

Next, Steven Folsom described a Cornell University pilot project involving posters and event flyers dating from 1977–1984 in the Cornell Hip Hop Collection. These ephemera contain a lot of information, including performers, venue, dress code, and date, and the collection already has substantial non-MARC metadata to draw on. Unlike Linked Jazz, which is built mainly on open linked data resources, this project was undertaken as part of the Mellon-funded Cornell/Harvard/Stanford Linked Data for Libraries (LD4L) project and explored the viability of using BIBFRAME 1.0 standards to convert existing library metadata to a linked-data-friendly format and connect it with other ontologies.

Descriptions of the collection were done in BIBFRAME, with other ontologies being used to augment or extend functionality in areas where BIBFRAME was still weak. Events and related entities were described using Music Ontology, Event Ontology, and Schema.org. Names were tracked using a combination of the FOAF (“friend of a friend”) schema and BIBFRAME authority data, retrieved from the LC Name Authority File, DBPedia, or local authority records. Additional data came from MusicBrainz, distilled through the Linked Brainz RDF mapping project.

As a proof-of-concept experiment, this project did not end with a complete publicly-visible interface, but did produce some valuable takeaways. BIBFRAME 1.0 was sufficient, though not ideal, for describing the collection, and feedback from this and other LD4L projects has been incorporated into BIBFRAME 2.0 development. It was also possible to map large parts of existing metadata into RDF, which then allowed more relationships to be brought in through interactions with other ontologies and databases. However, the workflow involved a good deal of manual preparation and processing and relied on software that was not always stable. Fully harnessing the relationship-exploring potential of the project would also require the existence of more URIs that could be linked to and linked back from.

A follow-up question from Kathy Glennan asked how the complex graphs produced by linked bibliographic data can be manipulated to produce usable visualizations. Folsom replied that UI does not necessarily need to be revolutionary; a more traditional catalog display can use fielded searches that are grounded in the underlying graph without necessarily using the graph as a UI.

Finally, James Soe Nyun discussed MLA involvement in linked data projects, beginning with those that produce *linkable* data: the NACO Music Project (formerly administered by the Music OCLC Users Group), the Medium of Performance Thesaurus, and the Genre/Form Task Force. The products of all of these initiatives are available via the Library of Congress’s Linked Data Service (id.loc.gov), which provides stable URIs for data and a basic visual exploration interface in addition to traditional narrower/broader term links.

The BIBFRAME Task Force has concluded its duties, which included testing MARC-to-BIBFRAME converters for medium of performance terms (still very much a work in progress), genre/form access points, and work access points. Work has also been done on developing BIBFRAME profiles—the forms one fills out while describing materials—for music. The successor will be a new BIBFRAME projects group, which will continue developing a Medium of Performance ontology, probably separate from but usable within BIBFRAME, and refining the profiles for music. Other tasks will include developing use cases for linked data in music and continued development of ontologies and MARC conversion tools. While the new group will need people familiar with BIBFRAME, input will be needed from others including archivists, reference and public services librarians, and catalogers; everyone is encouraged to get involved.