Linked Data Infrastructure Models: Areas of Focus for PCC Strategies

Prepared by members of the PCC Linked Data Advisory Committee: Jennifer Baxmeyer, Karen Coyle, Joanna Dyla, MJ Han, Steven Folsom, Phil Schreur, Tim Thompson
Date: June 2017

Introduction

The Program for Cooperative Cataloging (PCC) has a declared interest in exploring linked data as a method for sharing data (Program for Cooperative Cataloging, 2015). To better understand the requirements for making the transition from MARC to linked data, this white paper strives to describe unresolved areas of focus that the PCC should devote resources to as it seeks to establish viable linked data implementation and infrastructure models.

We begin by outlining the current metadata sharing landscape within the PCC. This is followed by anticipated challenges in transitioning to linked data workflows. We describe high-level functional requirements and the spectrum of models we expect to see as libraries and other cultural heritage institutions adopt linked data as a strategy for data sharing. Finally, we provide a set of takeaways and recommended next steps, along with an appendix describing related conversations.

We should be clear that a single collaborative model for distributing linked data has yet to emerge out of the very nascent experimentation taking place in libraries. Our hope is that this white paper will bring into focus areas of investigation for the PCC and the wider library community.

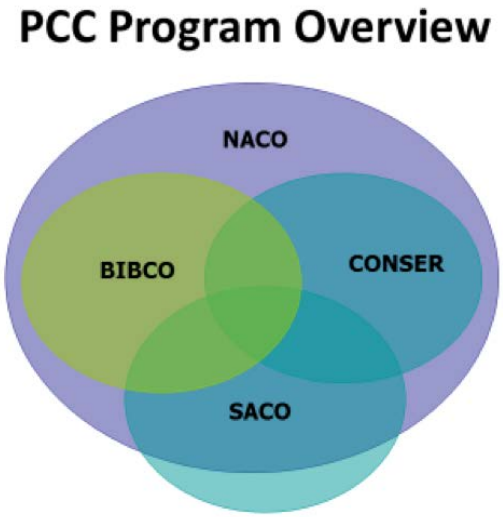

Figure 1. PCC NACO training, Module 1: NACO Foundations, Slide 15. Available at https://www.loc.gov/catworkshop/courses/naco-RDA/index.html.

How the PCC shares and manages data now

The PCC has four components (Figure 1; "Program for Cooperative Cataloging," n.d.):

- BIBCO (Monographic Bibliographic Record Cooperative Program), through which members contribute bibliographic records for a variety of formats, excluding serials, to globally distributed databases.
- CONSER (Cooperative Online Serials Program), through which members contribute bibliographic records for serials and integrating resources.
- NACO (Name Authority Cooperative Program), through which participants contribute authority records for personal, corporate, and jurisdictional names; uniform titles; and series headings to the LC/NACO Authority File (NAF).
- SACO (Subject Authority Cooperative Program), through which participants propose subject headings for inclusion in Library of Congress Subject Headings (LCSH) and classification numbers for inclusion in Library of Congress Classification (LCC) schedules.

BIBCO/CONSER

BIBCO and CONSER participants contribute authenticated bibliographic records for resources in all formats either by creating them directly in bibliographic utilities such as OCLC or SkyRiver or by batch loading locally created records to the bibliographic utilities, which then make them

available to their member libraries through copy cataloging. Library of Congress bibliographic records are freely available for downloading via Z39.50 directly from LC's online catalog and the bibliographic utilities provide access to these records as part of membership. BIBCO records can also be obtained directly from other libraries via Z39.50, as well as from vendors and through contract cataloging services. Bibliographic record updates and corrections are performed directly in the bibliographic utilities by PCC participants.

NACO/SACO

OCLC and SkyRiver provide platforms for NACO participants to create new authority records as well as update existing authority records. SACO member institutions submit proposals for subject headings to be added to or changed in LCSH and for new or changed classification numbers in LCC (Library of Congress, n.d.-a). Individual libraries can search either the LC Authorities, LC's Linked Data Service (http://id.loc.gov), or OCLC/SkyRiver for authority records. LC offers libraries the ability to download authority records (MADS/RDF or SKOS/RDF only) in bulk from its linked data service. In addition, LC offers subscription services for weekly FTP distribution of authority records.

Data enrichment

Throughout PCC's history a number of committees and initiatives have been charged with enabling and coordinating large-scale enrichments of our shared data. The Standing Committee on Applications, previously known as the Standing Committee on Automation, is charged with identifying and addressing issues related to automated processes that support the Program. During the transition to RDA cataloging, RDA task groups steered a sequence of large-scale phased changes to shared record files (Library of Congress, n.d.-b). Because of the governance around batch additions to shared library records, PCC libraries often use their own staff resources and/or vendor services to make batch enhancements to local copies of the records as a service to the wider community.

Current PCC linked data investigations

In addition to maintaining established processes for sharing MARC records, the Program has also actively engaged linked data related investigations.

Partnership with ISNI

A pilot group of "early adopter" libraries is being formed to try out a proposed model for a partnership between ISNI and PCC, with the goal of making ISNI an alternative to traditional authority work through NACO. PoCo also plans to create two task groups: Identity Management in NACO and PCC ISNI Implementation ("PCC umbrella membership in ISNI," 2017).

PCC Task Group on URIs in MARC

The primary goal of the Task Group on URIs in MARC is to identify and address any immediate policy issues surrounding the use of Uniform Resource Identifiers (notably HTTP URIs) in MARC records that should be resolved before implementation proceeds on a large scale. In collaboration with the PCC Standing Committees, it is developing guidelines for including identifiers in MARC bibliographic and authority records. Further, in cooperation with vendors and other partners, it is developing a work plan for the implementation of URIs in the $0 subfield and other fields/subfields in member catalogs and PCC-affiliated utilities. In consultation with the MARC Advisory Committee, technologists versed in linked data best practices, and other stakeholders, the group will identify and prioritize any remaining issues concerning support for identifiers in the MARC format and initiate MARC proposals as appropriate.

PCC BIBFRAME Task Group

From its charge:

> An overall goal for forming the task group is to involve the PCC program in actively developing practices around the BIBFRAME Initiative and other linked data activities.
>
> Jointly mapping of PCC standard records to BIBFRAME is the initial sub-project for this group. The mapping is a key MARC to BIBFRAME enabling project that will have many consequences for operationalizing PCC metadata in a BIBFRAME environment. (Program for Cooperative Cataloging, 2016)

Known challenges with linked data

By understanding how the PCC shares MARC records now and enumerating challenges and unanswered questions as we move to linked data workflows, we can begin to understand some of the infrastructure needs and areas of focus to consider in order to make linked data a viable data sharing strategy for the PCC Program.

Complexity of our workflows

Our workflows are increasingly complicated. Functions such as copy cataloging can vary from institution to institution based on choice of ILS, vendor services, and local cataloging traditions. In addition, the development of the digital repository has created its own metadata workflows and our database of record has become split between it and the ILS. The inclusion of article search and continuous vendor enhancements, such as tables of contents or reviews, makes the maintenance of our metadata a dynamic and ever-evolving process. Even the concept of what

is part of our collection is less clear because we have the ability to include links to free content within our catalogs.

And as libraries make the gradual shift to linked data, discovery will be an area of particular focus. For the foreseeable future, other functions such as payments will still be supported by MARC within the ILS and so we will need to be able to link discovery metadata to a parallel operational record within the ILS. Our discovery environments will need to take data from both in order to present our patrons with a complete picture.

Metadata for commonly held resources will become available in an increasing number of options: MARC, BIBFRAME, Dublin Core, schema.org, CIDOC-CRM, and so on. Infrastructure will need to have data conversion at its core or be able to handle multiple data types simultaneously. Similarly, workflows will need to be equally as flexible as data flows evolve.

Multiple bibliographic models and URIs for the same thing

In our current working environments, making use of data created outside of the library system, whether in MARC21 or in another format, requires that data be brought into the local database, usually after the data has been modified so that it resembles the targeted database format. This is often called a “crosswalk.” Along with crosswalks, linked data uses a different model for combining heterogeneous data. All linked data uses triples as its underlying data structure, and the database technology must be able to ingest any data as triples. To support functions such as search and display across different metadata sources, linked data has two primary technologies:

1. It is recommended that, where possible, metadata schema make use of existing vocabularies rather than defining new terms with the same meaning. Linked data allows unlimited mixing of terms from different namespaces. As an example, many schema make use of the FOAF (Friend of a Friend) vocabulary for the description of personal names and contact information. Many vocabularies make use of selected elements from the Dublin Core metadata standard, which is defined as a linked data vocabulary. Where different metadata schemas make use of the same vocabulary elements they have an automatic linking on those points in their descriptions.
2. If it is not possible to reuse existing vocabulary terms, links can be created between different terms that allow them to be used in a mixed environment. Those links can be horizontal (terms that mean the same thing), or vertical (terms that are super- or subordinate to each other). One could declare the BIBFRAME bf:title to be equivalent to Dublin Core dcterms:title. In doing so, any application that makes use of dcterms:title can also make use of bf:title and the subproperties of bf:title. This method is similar to

the creation of crosswalks with two important differences: it exists in the definition of the vocabulary, not as a separate function; and it makes no changes to the data descriptions.

Vocabularies that are defined without these points of linking exist as silos, and are unable to take advantage of the shareability of linked data. These features need to be included in the design of a metadata schema because they require coordination and are difficult to implement as an afterthought. Because the web of data is constantly growing, however, adjustments must be made over time to allow linking to new data sources in the sharing community.

Varying models

Linked data is designed for an open environment, often on the web. Not all data is suitable to be linked data, either because of its content or because of its business case. For example, one would not want private banking or medical data to be on the web. There are degrees of openness, ranging from entirely private to entirely open, and these can be managed within the linked data model. We see two predominant variations today within our own bibliographic environment: BIBFRAME and OCLC's use of schema.org.

The BIBFRAME model is a wholesale conversion of bibliographic data to RDF. This includes data elements that remain as textual entries with no linking function. BIBFRAME data will make use of linked data technologies to support local systems design as well as any linking to external sources that may be used.

Schema.org is an RDF-based vocabulary that is designed for use within web pages, often in a variety of RDF called RDFa that is a "lite" version of RDF, and increasingly in the JSON-LD format, which has been endorsed by Google (2017). The pages themselves are in HTML; the schema.org data is metadata within those pages, not visible in the web display, that encodes selected data for machine processing. It is used by online sales sites, for example, to make product offers more accurately searchable. OCLC has also added schema.org coding to its bibliographic display, even as it continues to store and offer MARC data as its basic data format. (See the section at the bottom of an OCLC bibliographic display that is called "Links" to view this.) Adding schema.org to the web display does not change the back-end OCLC database because schema.org does not replace the original data—it enhances it. With this method it is possible to create linked data only for those elements that must link, leaving much of the data in its original form.

The advantages of the total conversion model are that the data exists in a uniform linked data technology environment, using database and search options designed for that data. The

advantage of the RDFa model is that it is not necessary to convert one's data store or to define a linked data structure for aspects of the data that cannot be reasonably used for linking. In either case, one can decide to expose only the segment of its data that can be made public.

While it may be ideal that all metadata communities use the same vocabulary for the same things, it has always been clear that there will be different terms and different identifiers for the same thing. The difficult part is determining when two entities or terms are actually the same and when they have subtle but significant differences in meaning that can affect services. One solution to this is to leave the subtle differences in the local view of the data, and to expose a less nuanced view to the larger world. Thus a library database will have many different types of titles (title proper, work title, parallel title, key title), while a wider sharing environment may function with a single type of title such as Dublin Core title.

However, it is probably safe to say that the community that currently shares MARC data through structures such as BIBCO will need to agree on a single data vocabulary or a very small number of vocabularies that are designed to interact well with each other. Where changes will take place is in sharing with other partners, such as book vendors, publishers, and indexing services. These already use different metadata formats that must be integrated into library discovery systems. If one has been in the library field for a few decades the concept of federated search may be familiar, in which search engines translate a single search into many queries, each adapted for its target database, then provided multiple result sets. This tends to be awkward and not very user friendly. Linked data will make this kind of sharing a bit easier to develop, but will not overcome the differences between communities.

What do we need to share?

Entities and authority data

In current cataloging there is a separation between authority data and bibliographic data. Authority records are created that assign an authoritative label for an entity. That same string is entered into bibliographic data. Due to a lack of best practices around the use of URIs and other types of identifiers in metadata records, there is often not a direct machine-actionable link between the string in the authority record and the string in the bibliographic record.

In an entity-relation or linked data environment, an entity description is linked directly to any bibliographic description that refers to that entity. These entity descriptions may be fuller than today's authority record data because they are not limited to defining authoritative strings. In addition, they will be directly linked to the bibliographic description to provide information about the linked entities. Therefore, one set of data plays both roles: bibliographic description and

authoritative description of entities. We need to be sure that our linked data environments thoughtfully display information from different entity descriptions in ways library systems have struggled to do in the bibliographic/authority record model to date.

Bibliographic data

In a linked data context, there is much less of a division between the description of a primary bibliographic item and the description of those entities that are currently covered by authority records. All entities are in a sense equal, and all are within the purview of description. In FRBR-inspired models such as BIBFRAME, work (and expression), instance (or manifestation), and item entities also need to be differentiated and uniquely identified, just as name, title, and subject headings are in current practice. Traditionally, catalogers have created authority records for title strings (name-titles, uniform titles, conventional collective titles, and so on) only when needed for collocation and differentiation in alphabetically ordered browse-by lists. Now, in both BIBFRAME and RDA, descriptions of bibliographic entities are meant to be shared and potentially linked to, and so need to be addressable (although studies have suggested that the collocation and user experience improvements promised by FRBR may have been overly optimistic [Coyle, 2016, p. 114]).

Programs like the PCC could play an active role in standardizing the so-called entification process (Wallis, 2013) that linked data necessitates. One approach to the creation of entity URIs involves calculating unique fingerprints or hashes from key/value pairs that hold descriptive information about an entity (a specific application of hashing functions will be discussed below). Both the Library of Congress (Network Development and MARC Standards Office, 2016) and the data services vendor Zepheira have experimented with this approach, which Zepheira (2015) has spelled out in detail. However, in order for reconciliation based on hash values to be successful, the structure and content of the data to be hashed must be clearly defined. Moving forward, standards bodies may need to step into new roles and rethink the scope of their standardization activities—from creating rules for string curation to specifying algorithms for entification.

Libraries and other cultural heritage institutions will need guidance in constructing and applying new application profiles for bibliographic descriptions based on linked data. Defining a linked data equivalent to the BIBCO Standard Record, for example, may be a necessary task, although not an easy one. Combining and reconciling terms from different ontologies is a complex and often contentious process, especially when the needs of different domain communities and resource types must be accommodated.

Shifting to a graph-based, recordless paradigm removes the tidy frame that catalogers have been used to experiencing in current MARC-based workflows. In the semantic web community, distinct or circumscribed views of RDF graph segments are fluid constructs often referred to as "shapes." Work on methods for evaluating and validating RDF shapes is currently underway and has begun to coalesce around emerging standards such as the Shape Expressions Language (ShEx) (Prud'hommeaux et al., 2017) and the Shapes Constraint Language (SHACL) (Knublauch et al., 2017).

Where and how to publish linked data?

There remain many unanswered questions about where and how libraries and other cultural heritage institutions will publish linked data. The emerging linked data ecosystem relies on a range of protocols, specifications, and systems that may not be familiar to information technology staff in libraries. Although linked data allows libraries to publish their data directly to the web from their own local servers, and some will, many libraries will not have the expertise or capacity to implement linked data without using services hosted by vendors. Vendors typically have control over the domain name of the URIs in a hosted solution, and without the proper foresight and contractual agreements, vendor-provided linked data services may entail a certain amount of relinquished control over the data that an institution creates. Ideally the URIs created by third-party solutions would be vendor neutral (without reference to the vendor or platform providing access to the data). URIs should be owned by the libraries creating the data, and libraries should be able to migrate their data if they decide to use a different platform. A fallback and less desireable agreement might be that a vendor commits to support the redirection of traffic to library-defined replacement URIs.

Centralized versus distributed?

Centralized models

The transition to linked data shines a spotlight on some of the library profession's underlying assumptions about bibliographic control and quality assurance. To date, technical services workflows have relied on models that are primarily centralized in nature, where shared data is stored and maintained at a small number of locations on the web. The centralized model allows for gatekeeping: we control who is producing bibliographic data within our trusted spaces, and we can enforce community norms and standards. The centralized model also simplifies where to focus data enhancements so that the greatest number of collaborators can benefit; adding improvements to a few known locations assures that those benefits will reach the widest audience. However, with centralization, only a relatively small number of contributors are

permitted to provide both manual and batch level value-added enhancements, and the potential range of partners is arbitrarily limited.

In addition to limiting participation and opening the door to vendor lock-in, there are other drawbacks to centralization, and the distributed nature of linked data makes the fragility of the centralized model more apparent. When central systems and servers are relied upon in order to accomplish mission-critical tasks, the degree of risk increases because there is a single point of failure. In a linked data environment, compared to the current MARC-based environment, the number of data points will grow exponentially, and the need for high availability of data and services will be of paramount importance.

One common misconception in libraries is that linked data will afford us the ability to avoid making local copies of data created elsewhere. Given the current state of federated searching and the complications it presents ("Federated searching," 2017) during the near future it is likely any data libraries want to query within local systems will require local indexing. Systems might consider fetching data in real-time if the data is being used only for display, but even this requires confidence that the data source is performant (meeting demand at an acceptable speed) and will remain available at the time of need. Other motivations for storing data locally include the ability to assert some level of provenance and control over it: for example, to assure accurate inventory or to protect certain data behind firewalls (the latter considered linked enterprise data, rather than linked *open* data [Crandall et al., 2013]).

Whether copying data locally or not, in order to consume existing linked data we need to understand how the data is made available for consumption. Because linked data/RDF can be offered in any number of ways (for instance, dereferenceable URIs, RDF dumps, SPARQL endpoints, RDFa, Linked Data Fragments), data consumers are challenged to use many methods of acquiring data. Often the methods available for a particular desired dataset do not meet all the needs of a consumer. An example of this would be if the consumer did not have a URI for a resource believed to be described in a dataset that only made its data available through dereferenceable URIs; without the URI for an entity it is difficult for the consumer to access the data.

The systems and data storage requirements involved in producing and publishing linked data inevitably bring us back to the gaps in expertise and capacity mentioned above. Without prior implementation experience to draw upon, libraries may find themselves increasingly reliant on third-party services for data conversion, creation, and hosting. Ultimately, libraries will need to perform a cost/benefit analysis, comparing the cost of vendor-based solutions to the cost of investments in local infrastructure and capacity building. With limited funds to allocate, technical services administrators may continue to find centralized approaches attractive, especially during

the transitional period in which hybrid MARC/linked data workflows will predominate. Compared to distributed approaches, centralization is more efficient and easier to manage: having a single point of failure also means having a single, identifiable point of contact for technical support, rather than a dispersed community of peers. Again, the risk involved here is that libraries may become dependent on vendor services from the outset, making any future transition to greater autonomy in data management and ownership more difficult.

A focus on the role of vendors does not tell the whole story, however. There are alternative approaches to centralization that attempt to leverage its benefits while maintaining an emphasis on openness and collaboration. Perhaps the foremost example of this approach is Wikidata (https://wikidata.org/), which, like Wikipedia, is a platform hosted by the Wikimedia Foundation. Wikidata is a centralized knowledge base that serves structured data to all other Wikimedia projects, including Wikipedia, Wikisource, and Wikibooks. Additionally, Wikidata maintains a public SPARQL endpoint and serves its data in a variety of RDF serializations. As a wiki itself, Wikidata's content can be collaboratively created, edited, or deleted. Each modification is recorded in an item's history, and specific versions can be referenced or reverted to if necessary.

In the context of Wikipedia, a hub like Wikidata makes it possible to transcend language barriers and efficiently synchronize data updates. There are currently 285 active Wikipedias, which correspond to distinct language communities (such as the English Wikipedia, Spanish Wikipedia, Arabic Wikipedia, and so on). Before the advent of Wikidata, edits to structured data elements across wikis needed to be made manually or through the creation of a bot that could make automated updates. Now pages hosted on different Wikipedia instances, but devoted to the same topic, can reference the same Wikidata URIs and receive updates automatically. As libraries move forward with linked data implementation, possibilities for collaboration with open knowledge initiatives such as Wikidata should be encouraged and actively explored.

Distributed models

Paul Baran's 1964 publication, *On distributed communications: introduction to distributed communication networks* (prepared for the United States Air Force Project RAND), describes network models in terms of their vulnerabilities. Baran's now famous diagram characterizing centralized, decentralized, and distributed models is useful to illustrate the spectrum of options library communities have at their disposal for communicating about their collections and sharing data.

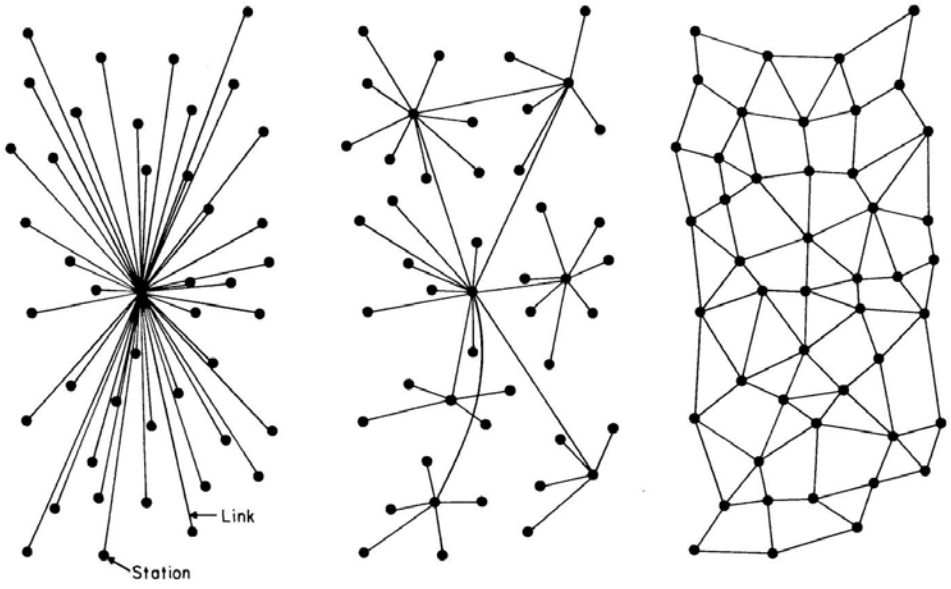

FIG. 1 — Centralized, Decentralized and Distributed Networks

As alluded to previously, current library data flows tend to follow more centralized models, given our heavy reliance on OCLC, SkyRiver, the Library of Congress, and other national libraries for sharing data. Linked data allows us to consider models closer to the distributed end of the spectrum, which as Baran points out, is the least vulnerable to large scale outages. In the context of linked library data, distributed models also give us the ability and agency to experiment with saying things locally while still making the data available globally for others to use.

Distributed, peer-to-peer systems first gained widespread attention with the rise of file-sharing services such as Napster, Kazaa, and BitTorrent. Since then, decentralized models have evolved significantly, particularly with the emergence of the digital payment network Bitcoin (https://bitcoin.org), a “cryptocurrency” that relies on sophisticated hashing algorithms and distributed methods in order to process transactions and store data. Bitcoin was designed to address many of the shortcomings of centralized financial markets, which are vulnerable to fraud and require users to divulge personal identifying information. It is a “trustless” system in that transactions are verified algorithmically through a complex process called blockchain, which provides privacy safeguards and reduces the potential for fraud, manipulation, or identity theft (Driscoll, 2015).

An emerging technology that is more directly relevant to linked data infrastructure for libraries is the InterPlanetary File System (IPFS), which is based in part on technologies employed by Bitcoin. IPFS embodies a new vision of the World Wide Web: rather than addressing content by location, it implements a method known as content addressing (Benet, n.d.). On the current web, URLs point to specific locations on specific servers so that clients can retrieve content. In IPFS, by contrast, content is referenced using one-way cryptographic hashing algorithms that uniquely identify each data object (Zumwalt, 2017). The “InterPlanetary” in IPFS refers to its globally distributed nature. Dereferencing an IPFS hash URI allows clients to retrieve content from anywhere, independent of its physical location. Data is replicated across many geographically disparate nodes, minimizing the chance that it will cease to be available.

IPFS also offers a solution to the problem of online digital preservation and so-called link rot and makes it possible to verify the authenticity and integrity of a resource using digital signatures. The unique identifier generated from an object stored on IPFS captures its state at a specific point in time, providing a form of version control. IPFS identifiers are permanent because an object's identifier, which is a hash representation of its content, is guaranteed not to change. The technology behind IPFS has the potential to provide a robust infrastructure for linked data applications and storage, making the web of data more persistent and resilient.[1]

In the end, simply because a system is distributed does not mean that it constitutes a free-for-all. The use of Bitcoin, for example, is governed by specific rules and procedures that all participants in the market are expected to abide by. Similarly, as libraries contemplate the implications of embracing the *open* in linked open data, they need not abandon the rules and standards that have been created to ensure the consistency and quality of their data. If anything, those rules and standards become even more important in a distributed system, which can only function if its stakeholders agree to adhere to a common set of norms.

Sharing and notifications

In addition to IPFS, other distributed approaches have recently emerged from within the linked data community. A leading role in this regard has been taken by the Social Web Working Group (Guy, 2017), whose members have developed protocols, specifications, and software to enable peer-to-peer interaction using linked data. This interaction is enabled through authentication methods such as WebID (WebID Incubator Group, 2013), which replaces traditional passwords with URIs that are associated with unique identities; these individual identities, in turn, can be verified using private/public cryptographic key pairs.

[1] The data model underlying IPFS is called IPLD (InterPlanetary Linked Data). The term “linked data” has a particular meaning in the context of IPFS, but many of the concepts are compatible with the RDF model.

One social web protocol of particular relevance to libraries is the Linked Data Notifications (LDN) protocol. LDN defines a mechanism for data sharing that allows users to create Inboxes and send and receive notifications. As the LDN specification describes it:

> Linked Data Notifications (LDN) supports sharing and reuse of notifications *across* applications, regardless of how they were generated. This allows for more modular systems, which decouple data storage from the applications which display or otherwise make use of the data. The protocol is intended to allow senders, receivers and consumers of notifications, which are independently implemented and run on different technology stacks, to seamlessly work together, contributing to decentralisation of our interactions on the Web. (Capadisli & Guy, 2017)

LDN and other social web technologies are now being implemented in front-end clients such as dokieli (https://dokie.li/) and projects such as Solid (https://solid.mit.edu/), which stands for "social linked data." In the context of library metadata, one can easily imagine the role that LDN could play in cooperative cataloging: catalogers could define trusted sources to receive notifications from whenever metadata is created or updated for a resource that is held in common.

Social web technologies built on linked data are still in the early stages of development and implementation and are primarily focused on interactions between individuals rather than institutions. The library community could make a contribution to the development of these technologies by extending their scope to support large-scale data sharing and updates on an institutional level.

Governance

Moving to a culture of greater data sharing means that each community must look beyond its own silo when developing standards. Rather than developing a standard that works for only one group, such as libraries, it becomes important to take into account one's desired sharing partners. This means collaborating with other communities. A trend that has begun, that might be accelerated with the move away from MARC, is greater collaboration among different sections of the GLAM (Galleries, Libraries, Archives, and Museums) community. Expanding the scope of collaboration even further, we can look for synergies with the semantic web community. An early example of this is the collaboration between OCLC's VIAF project and Wikipedia. The addition of VIAF identifiers to Wikipedia pages makes it possible to connect Wikipedia readers with library holdings, and links to Wikipedia can take library users to further information about persons and works.

This emphasis on sharing across communities also means that the standards process must be open to a broader constituency and that standards must be discoverable and comprehensible to that larger group. We will need to become more adept at explaining our practices so that others can find ways to collaborate with us. One good way to foster wide collaboration is to develop standards in general standards bodies, such as the World Wide Web Consortium and the International Standards Organization. These standards cross the boundaries of individual communities, and are the go-to places for technologists looking for accepted international standards.

Costs

The costs connected with the transition to RDF for data creation, management, and discovery are projected to be high and will require an institutional-level commitment. Cultural heritage institutions will need to invest in staff training to support new skill sets; linked data fundamentals will be required by at least some staff, but new cataloging interfaces can more closely reflect cataloging rules, providing a more guided experience for cataloging professionals than they are typically used to in current cataloging interfaces. Libraries will need to invest in vendor/developer relationships to advocate for functional requirements for linked data platforms. Cultural heritage institutions do not always consider the investment in standards and community development an operating cost, but as staff gain the fundamentals and experience it critical that they be encouraged to share their experience outside of their local institutions to accelerate the adoption of linked data.

Many fundamental services such as the ILS are rooted in MARC. Libraries create and exchange data in MARC. Vendors supply data in MARC or perform services built on top of MARC data. This entire support structure will need not only to be modified but redefined. Libraries and vendors will need to support their current systems and workflows as they simultaneously support the development of new ones. As the community defines what it means to perform functions such as original or copy cataloging in a distributed environment, undoubtedly efficiencies will be found. However, additional functions such as reconciliation will need to be developed. These trade-offs together with advances in automation will most likely keep costs stable after the transition period. The return on investment should be defined in terms of improved discovery and better integration with the web rather than an expectation that linked data might provide enough efficiencies to justify cuts in technical services spending.

High-level functional requirements

Like many technologies, linked data is flexible in its application and can be designed and utilized to address a wide variety of functions. For this reason, at this point in time there is much about

our bibliographic future that has yet to be determined. This early stage of experimentation with linked data is the time in which we need to explore a wide range of possible scenarios and functions.

Ideally the library community, with the cataloging community in a key position, will develop a series of questions that can stimulate thinking about desired design principles. These questions should cover the full range of functions that make use of the catalog data and its interfaces. Some examples of the types of questions that are needed are:

1. What is the user view of library materials that is desired? Is it the same for all material types? Is it the same for each library?
1. Are there internal bibliographic links (work/work, expression/expression, manifestation/manifestation) that we want to use, either for searching or for display?
2. What links do we want to be actionable in our catalogs? For example, do we wish to link between library catalogs and online resources like Wikipedia, GeoNames, and MusicBrainz?
3. What entity information is needed during the cataloging process to help catalogers make correct decisions as efficiently as possible?

A number of design methodologies exist that are used by communities undergoing important technological changes. They vary in their details but the general substance is that they all encourage taking a broad look at the needs of the community and the technologies that can answer those needs.

This section addresses only a small number of the questions that need to be answered, but could be used as a starting point for a wide-ranging community discussion.

When to use local versus external URIs

Nearly all computer systems assign and make use of internal identifiers that support functions like storage and retrieval. These internal numbers are rarely visible to human users of the systems, and have no meaning outside of the local systems. We are more aware of public identifiers, like ISBNs or passport identity numbers. These may be seen on items in an eye-readable form even though they are also stored in systems that make use of that identity.

Similarly in systems designed to create linked open data we employ a combination of local URIs (URIs within namespaces under an institution's control) and linking to external URIs (URIs controlled by others in their own namespace). These URIs can be used locally to serve similar functions as internal system numbers, but they are also useful as global dereferenceable URIs, allowing external systems (through HTTP and other API protocols) to retrieve and use the same

data. Under no circumstances should humans be keying such identifiers into metadata. These URIs should be clickable and should show the user helpful information about the thing being identified.

Ideally, one would link to existing URIs and only create a local URI for an entity if an institution has something new to say about the entity and does not have access to make new assertions directly on the existing entity. Practically speaking, however, we know that multiple URIs for the same entity will be created for any number of reasons, for example, institutional policies to create and maintain all data locally, the inability to discover existing entities during different moments in workflows, local system requirements, different points of view about the entity. In these cases where new URIs are created when other URIs for an entity already exist, we ultimately should strive to link the two URIs to enable the most complete view of the entity. It is in this setting that highly visible "hub" datasets (id.loc.gov, Wikidata, VIAF, ISNI, FAST, and so on) play an important role in making connections on the semantic web. Libraries have the ability to contribute to and link to these hub data sources, and linked data allows libraries the freedom to make more and different assertions locally while still connecting to the larger linked data ecosystem.

Conversion of non-RDF data

We anticipate that libraries will create catalog records in the MARC format for the foreseeable future because library systems that support linked data from the point of creation are very nascent, and many of our existing metadata workflows are not set up to accept linked data, continuing instead to require MARC and other non-RDF formats. Then the question arises: how does a library convert non-RDF data to linked data? As native linked data cataloging tools emerge, how do we convert back to legacy formats for those collaborative cataloging workflows not yet ready to consume linked data? The actual conversion can be done relatively easily by using the conversion tools already published by the Library of Congress and others, using shared or locally created customized mappings to specific ontologies. However, libraries still need to design workflows for the points at which reconciliation (with which linked data sources?) and conversion (to which linked data ontologies?) should happen. Thought needs to be given to how to reflect changes made in MARC records to the previously reconciled and converted RDF data, that is, whether the one system will support both non-RDF and RDF data, and if not, how the two systems will work together in a conversion workflow.

Many library linked data conversion efforts focus on traditional MARC library data, leaving a lot of questions unanswered about how to convert other library data, such as digital collections, archival resources, and research data. If all parts of library collections are to benefit from linked data treatment, converters need to be flexible enough to accept and understand data in various file formats and schemas, and recognize data created according to different practices. Several

generic RDF converters have been created by the semantic web community, and these tools should be evaluated to see if they are flexible enough to meet library needs.

Discovery, linking, and caching of data

With libraries and other cultural heritage institutions likely publishing data to various data hubs and potentially asserting information in local systems (on local servers), there will be a need for caching or distributed storage of the data both to ensure its availability and to provision for common data discovery services. Central discovery of existing entities will allow a library to discover, link to, and take advantage of existing data on the web. As services for uniform discovery of data emerge, these services will need to define their selection criteria and allow data providers the ability to register their datasets. Models offering centralized discovery of data do not preclude more peer-to-peer discovery of data, whereby institutions can build targeted workflows to take advantage of focused connections between specific datasets.

As stated above, existing centralized record-based workflows provide obvious destinations for batch enhancements to community managed library data. These enhancements also require a significant amount of coordination and currently only a few are permitted to make batch changes to the central database of record. Through more distributed creation and maintenance of data on the web, we leave open the option that any institution or set of institutions can contribute high-value enhancements to the semantic web graph, and through centralized discovery of the data libraries ought to be able to decide what part of this data to take advantage of locally.

RDF editors (linked data cataloging tools)

It is fair to say that a large amount of work remains to be done in developing RDF editors for the library community. Linked data cataloging tools often focus on and reflect the underlying data models, which are not always how catalogers conceive of their descriptive practices, nor do the models necessarily map to cataloging workflows. User interactions with these cataloging tools need to expose the complexity of the data models only to staff who need to evaluate the data for semantic and structural coherence (similar to how current MARC cataloging tools do not require all users to see the underlying MARC). As mentioned above, discovery of existing data is critical in nearly every step in cataloging workflows; RDF editors, in order to meet the goals of linked data in libraries, need not only to be able to create new RDF from scratch, but also to recognize and link to existing data. If the library creating data needs to assert more information about the external entity, they can either assert this information on a local URI (preferably including a link between the local and external entities), or if they have the appropriate permissions, the library can assert this information directly on the existing external entity.

Dataflows and APIs

In order to provision for discovery and access of data (as described above), there needs to be a greater focus on dataflows and the APIs (Application Programming Interfaces) that allow interchange between systems. These interfaces need to be stable and performant and to provide the greatest possible degree of access to the data itself; often APIs provide a narrow view into a particular dataset, whereas the ideal linked data scenario is to provide as much of the data to consumers as possible in order to enable both anticipated and unanticipated uses. It is also imperative that metadata about changes, provenance, and data quality be built into these dataflows. Ultimately it is possible that we might use data provenance and RDF-based methods (like SHACL and ShEx) to define and test for measures of quality, rather than authenticated centralized data repositories, to assure that the data we consume meets library standards.

With the primary goal of connecting data across the web, assuming agreements are made between collaborators, APIs should not only provide read access to data, but also the ability to write to others datasets. Short of directly writing to others datasets, we need the ability to notify data providers when we have made an assertion about an entity that they also have data about, essentially a giving the other data provider a chance to create a back link and take advantage of the new assertions we have made. Again, libraries might consider experimenting with Linked Data Notifications (LDN) as a method for notifying and being notified of statements made about entities of shared interest.

Ontology management and mappings

The library community has not settled on a single bibliographic ontology (a lingua franca) to describe our collections. There continue to be attempts to achieve consensus around a single bibliographic ontology, which makes a lot of sense for interoperability reasons, but it is unlikely that all institutions will use the same model to describe their collections. Many libraries will implement BIBFRAME, but it will not stop others from using schema.org, dcterms, BIBO, RDA, etc. The challenge for a shared model is further complicated when considering non-bibliographic entities like persons, places, events, concepts that provide context to our collections. Even if all PCC participants agree to use the same model or models, we would not want to rule out linking and consuming data from non-PCC collaborators. Acknowledging this reality means one of the functional requirements of a linked data infrastructure is the ability to manage and map data expressed in various models. Like MARC and other metadata standards, ontologies evolve over time, so systems will need to be able to understand different ontology versions and handle data accordingly. It is possible, depending on how the ontologies are managed, that cataloging systems may also need the ability to create new classes and properties to extend existing models.

Reconciliation of instance data

With libraries and other cultural heritage institutions potentially asserting information locally about entities that are already described on the semantic web (within the library community and also the wider semantic web), there will be a need for reconciliation of multiple URIs for the same entity. The algorithms used to identify potential matches need to be transparent and adjustable to account for different workflows and metadata practices. Experience in this area suggests that successful matching requires a balance of automated and cataloger supervised workflows. The need for reconciliation will be compounded as each library converts its local catalog to linked data, otherwise the result will be local URIs for the same entities across library linked data implementations.

Conclusion: takeaways and next steps

With so many unknowns and so many linked data implementation models open for exploration (see Appendix A), a shared infrastructure for the PCC to distribute linked data is likely going to consist of a heterogeneous mix of different platforms and services. Eventually there may be convergence around particular platforms and services, but it is too early to predict.

What we do know is:

- Library linked data implementation models/services need to integrate with our MARC workflows during this transition period and perhaps longer.
    - We need production-ready MARC to RDF converters for as long as MARC encodes our database of record and/or data providers are limited to creating MARC.
    - For the short term, libraries will likely want an RDF to MARC converter to fulfill MARC record distribution needs and ILS inventory functions.
- Libraries require RDF editors optimized for describing library collections and flexible enough to allow libraries to make their own choices about what information to capture and publish.
- Libraries will need ways to discover (and ensure “uptime” for) the linked data it wants to consume, likely through a combination of coordinated caches/mirrors and linked data discovery services supporting dynamic look-ups.
- More focus needs to be given to establishing dataflows and APIs, addressing issues of how data will be passed from one system to another, updates promulgated, and provenance preserved.

- Libraries need more experience evaluating linked datasets and associated techniques, defining what quality means in a linked data context.
- Libraries are unlikely to agree to use the same ontologies to create and consume data about their collections for the foreseeable future, necessitating infrastructure choices that support ontology management, instance data migration to new/different versions of the same model, and mappings between different models.
- Given the inevitability of multiple URIs for the same entity, reconciliation services are critical for making connections between different URIs for the same thing.

In terms of next steps, PCC and its membership need to engage with vendors and open source communities to articulate a set of functional requirements for linked data tools and infrastructure (the requirements briefly described above and others as they become apparent). Vendors have begun to devote significant resources toward building linked data tools, but they will be reluctant to continue to invest in this area if they do not see a demand or have clear direction regarding which services to provide. In the same way, the datasets we want to link to and consume must be identified so that specific dataflows can be engineered. Again, we should not artificially limit the scope or nature of our collaborations, but rather expand the radius of our partnerships beyond our traditional library circles to GLAM partners and the wider linked data community. This will expose us to prior art (existing tools, datasets, and so on) that we can evaluate, reuse, and extend to meet our bibliographic description needs. Through targeted participation in standards development in the linked data community, the PCC can ensure that the approaches we take and the data we produce will benefit from expertise and economies of scale much greater than what the library community alone can provide.

References

Baran, P. (1964). *On distributed communications: introduction to distributed communication networks*. Santa Monica: Rand Corporation. Retrieved from http://www.rand.org/content/dam/rand/pubs/research_memoranda/2006/RM3420.pdf

Benet, J. (n.d.). IPFS—content addressed, versioned, P2P file system (DRAFT 3) [whitepaper]. Retrieved from https://ipfs.io/ipfs/QmR7GSQM93Cx5eAg6a6yRzNde1FQv7uL6X1o4k7zrJa3LX/ipfs.draft3.pdf

Capadisli, S., & Guy, A. (Eds.). (2017, May 2). Linked Data Notifications: W3C recommendation 2 May 2017. Retrieved from https://www.w3.org/TR/ldn/

Coyle, K. (2016). *FRBR: before and after*. Chicago: ALA Editions.

Crandall, M., et al. (2013). Planning a platform for learning linked data. In *Proceedings of the International Conference on Dublin Core and Metadata Applications 2013*. Retrieved from http://dcpapers.dublincore.org/pubs/article/download/3693/1916

Driscoll, S. (2016, June 15). Introduction to Bitcoin and decentralized technology. In *Pluralsight* [online learning platform]. Retrieved from https://app.pluralsight.com/library/courses/bitcoin-decentralized-technology

Ennis, M. (2016). Library.Link builds web visibility. *Library Journal, 141*(13), 18-19.

Federated searching: challenges. (2017, March 10). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Federated_search#Challenges

Google. (2017, May 26). Introduction to structured data. Retrieved from https://developers.google.com/search/docs/guides/intro-structured-data

Guy, A. (Ed.). (2017, May 4). Social web protocols. Retrieved from https://www.w3.org/TR/social-web-protocols/

Institute of Library and Museum Services. (2016). LG-73-16-0040-16. Retrieved from https://www.imls.gov/grants/awarded/LG-73-16-0040-16

Knublauch, H., et al. (Eds.). (2017, June 8). *Shapes Constraint Language (SHACL): W3C proposed recommendation 08 June 2017*. Retrieved from https://www.w3.org/TR/shacl/

Library of Congress. (n.d.-a). About the SACO program. Retrieved from https://www.loc.gov/aba/pcc/saco/about.html

Library of Congress. (n.d.-b). RDA Task Groups. Retrieved from https://www.loc.gov/aba/pcc/rda/RDA%20Task%20Groups.html

Network Development and MARC Standards Office. (2016, October 14). *marc2bibframe* [GitHub repository]. Retrieved from https://github.com/lcnetdev/marc2bibframe

PCC umbrella membership in ISNI: next steps after PoCo endorsement. (2017). Retrieved from https://docs.google.com/document/d/1pEZYl9AVt7j-iJA2hvTeNnaYHvgew25v634sZ6EPpco/edit

Program for Cooperative Cataloging. (2015, November 20). Vision, mission, and strategic directions: January 2015-December 2017. Retrieved from https://www.loc.gov/aba/pcc/about/PCC-Strategic-Plan-2015-2017.pdf.

Program for Cooperative Cataloging. (2016, September 20). PCC BIBFRAME Task Group (charge revised September 20, 2016). Retrieved from https://www.loc.gov/aba/pcc/documents/PCC-BF-TG-Charge.docx

ProQuest. (2017, May 8). Ex Libris increases library connectivity with implementation of BIBFRAME roadmap. Retrieved from http://www.proquest.com/about/news/2017/Ex-Libris-Increases-Library-Connectivity-with-Implementation-BIBFRAME.html

Prud'hommeaux, E., et al. (Eds.). (2017). *Shape Expressions Language 2.0: draft community group report 27 March 2017*. Retrieved from http://shex.io/shex-semantics/

University of Illinois at Urbana-Champaign. (2017). Linked open data for special collections. Retrieved from http://publish.illinois.edu/linkedspcollections/about/

Wallis, R. (2013, March 12). From records to a web of library data—pt1, entification. Retrieved from http://dataliberate.com/2013/03/12/from-records-to-a-web-of-library-data-pt1-entification/

WebID Incubator Group. (2013). WebID specifications. Retrieved from https://www.w3.org/2005/Incubator/webid/spec/

Zepheira (2015, September 10). From records to resources: the Library.Link resource ID generation algorithm. In *pybibframe* [GitHub repository]. Retrieved from https://github.com/zepheira/pybibframe/wiki/From-Records-to-Resources:-the-Library.Link-resource-ID-generation-algorithm

Zumwalt, M. (2017). *Decentralized web primer* [GitBook]. Retrieved from https://www.gitbook.com/book/flyingzumwalt/decentralized-web-primer/details

Appendix A

Related library conversations in this space

Please note the following list is not exhaustive.

Local Authorities IMLS Grant

Grant Description: Cornell University Library, in partnership with the Library of Congress, OCLC, the Program for Cooperative Cataloging, the ORCID organization, the Coalition for Networked Information, the Social Networks and Archival Context Cooperative, the BIBFLOW project, Stanford University Library and Harvard Library will hold a national forum on issues concerning local name authorities. Name authority files provide unique identifiers and records for people to ensure consistency in creation of descriptive metadata." Libraries create local authorities to serve a variety of purposes, usually within an institutional context; but these authorities have significant potential for reuse at other cultural heritage organizations and beyond. The April 2015 IMLS National Digital Platform Forum report emphasized the importance of enabling technologies (e.g., interoperability via linked data) and radical collaborations in supporting the mission of the cultural heritage sector. By facilitating a national forum, we plan to identify solutions for facilitating the creation of more shareable authorities (Institute of Library and Museum Services, 2016).

Digital Special Collections Mellon Grant

University of Illinois at Urbana-Champaign Library is experimenting Linked Open Data (LOD) for digital special collections. The project was funded by the Mellon Foundation in 2015 and is seeking to advance our understanding of the question “after digitization, what more needs to be done to maximize the usefulness of these digitized [special collection] resources?” (University of Illinois at Urbana-Champaign, 2017). While there are several notable LOD projects with general collections library metadata, projects examining LOD for special collections and considering how user interfaces to these collections might be enhanced by LOD are fewer. This project is focused on transforming legacy special collections item-level metadata into schema.org RDF and integrating LOD into services and end-user interfaces. Through two small user studies, the project is hoping to get a better sense of ways that LOD work might increase the visibility of special collections on the Web and usefulness of special collections for scholars. The project will wrap up by the end of 2017 and share the workflow scripts employed and findings from the user studies.

LD4P

Grant Description: Linked Data for Production (LD4P) is a collaboration between six institutions (Columbia, Cornell, Harvard,Library of Congress, Princeton, and Stanford University) to begin the transition of technical services production workflows to ones based in Linked Open Data (LOD). This first phase of the transition focuses on the development of the ability to produce metadata as LOD communally, the enhancement of the BIBFRAME ontology to encompass the multiple resource formats that academic libraries must process, and the engagement of the broader academic library community to ensure a sustainable and extensible environment. As its name implies, LD4P is focused on the immediate needs of metadata production such as ontology coverage and workflow transition. The LD4P partners' work will be based, in part, on a collection of tools that currently exist, such as those developed by the Library of Congress (used by LC and Stanford) and VitroLib (developed by Cornell and used by Cornell and Columbia). The cyclical feedback of use and enhancement request to the developers of these tools will allow for their enhancement based on use in an actual production environment. The focus of LD4P is on the adaption of this current tool suite to immediate production needs. The individual institutions involved in LD4P will be working on separate, but related, projects:

- Columbia: Development of a BIBFRAME extension for the description of two- and three-dimensional art objects
- Cornell: Development of a BIBFRAME extension for rare materials and BIBFRAME's use for the original cataloging of non-commercial LPs
- Harvard: Development of a BIBFRAME extension for the description of cartographic materials
- Library of Congress: BIBFRAME for the description of audiovisual and sound recordings, prints and photographs, revision of the BIBFRAME ontology (BIBFRAME 2.0), and BIBFRAME and Resource, Description and Access (RDA - our cataloging standard)
- Princeton: BIBFRAME and annotations in relationship to the library of Jacques Derrida
- Stanford University: Development of a BIBFRAME extension for performed music and the initial conversion of four workflows to LOD

LD4L-Labs

With the support of the Andrew W. Mellon Foundation, *Linked Data for Libraries: LD4L Labs* is a collaboration of Cornell, Harvard, Iowa, and Stanford to continue to advance the use and usefulness of linked data in libraries. Project team members will create and assemble tools, ontologies, services, and approaches that use linked data to improve the discovery, use, and

understanding of scholarly information resources. The goal is to pilot tools and services and to create solutions that can be implemented in production at research libraries within the next three to five years. Efforts focus on the enhancement of linked data creation and editing tools, exploration of linked data relationships and analysis of the graph to directly improve discovery, ontology development around BIBFRAME and piloting efforts in URI persistence, and metadata conversion tool development needed by LD4P and the broader library community. In addition to the LD4P institutional domain projects, within the LD4L project Harvard is creating BIBFRAME extension for the description of the Harvard Film Archive collection.

Vendors

The following list of vendors building linked data functionality in their offerings is not an endorsement, but, rather, it is meant to serve as a sample of the increasing number of vendors working in this space. Again, this is not an exhaustive list; the vendors creating linked data tools and what these tools are attempting to achieve is constantly evolving. For complete information about what linked data services a particular vendor offers or is scheduled to offer, please consult with vendors directly.

Casalini SHARE-VDE

Through the processes of analysis, enrichment, conversion and publication of metadata from MARC21 to RDF and considering libraries various systems and needs, Casalini will develop a new discovery environment called SHARE-VDE based on BIBFRAME's three layer architecture (Work/Instance/Item). In addition, the project will create a database of relationships to make evident the relationships between elements contained but not expressed in the MARC data. Through an iterative process of conversion and evaluation of data in the discovery interface, Casalini will evolve the service to meet library needs and configure the services as individual subscription elements. The current phase of the project has twelve members and may be expanded in future. A complete prototype will be demonstrated at ALA Annual 2017. The project website has started a new section on specific library use cases. Initial discussion have focused on interlibrary loan and the use of the RDF data for "copy-cataloging" purposes.

Ex Libris

Ex Libris and the ELUNA/IGeLU Linked Open Data Special Interest Working Group has been experimenting with linked data since at least 2011; the linked data support implemented by Ex Libris in 2016 is a culmination of this experimentation. As Ex Libris carries out its linked data and BIBFRAME roadmap, it is focusing first on a MARC to BIBFRAME converter allowing for publication of BIBFRAME using the ALMA library service platform. Beginning with conversion is an attempt to allow clients to produce linked data while minimizing disruption of current

workflows. The Ex Libris roadmap includes the eventual creation of a BIBFRAME cataloging tool (ProQuest, 2017).

OCLC

Initiatives to take data into the future: Based on linked data research and pilot programs, OCLC is exploring ways to embed linked data relationships into WorldCat to ensure that user searches deliver rich and relevant results. In addition, OCLC is studying ways to improve their services around VIAF, ISNI, and WorldCat Works and Persons to enable new models of cooperation and description.

WorldCat works: In April 2014, OCLC released 197 million WorldCat work entities that bring together multiple manifestations of a resource into one authoritative record. As of May 2017, more than 215 million WorldCat work entities are available. WorldCat work entities connect all descriptions of a work, despite variations in titles, publishers, authors' names, subject headings and other bibliographic information.

WorldCat persons: WorldCat person entities connect related information about specific people into a brief description that includes various formats of the person's name, creative works that the person has produced, and biographic sources of information about the person. As of May 2017, WorldCat persons include more than 117 million descriptions of authors, directors, musicians and others, which have been mined directly from WorldCat. In 2016, OCLC conducted a linked data pilot program in which libraries used WorldCat persons in their regular workflows.

Linked data as a cooperative effort: OCLC works closely with other organizations, such as the Library of Congress, W3C and other data standards groups, to participate in linked data discussions and initiatives, ensuring that library data are included on the web.

Zepheira

Zepheira (https://zepheira.com), a linked data consulting company, initiated several library linked data experimentations and services. Starting in 2014, Zepheira has been providing a linked data training program for information specialists and initiating linked data projects including Library Hub and Library Link. Library Link covers the lifecycle of linked data, including conversion to BIBFRAME, reconciliation entities with linked data sources, publishing data to the web with a creative commons attribution, and consuming linked data resources for user services. Zepheira has been also working with vendors, including EBSCO, Atlas Systems, Innovative , and SirsiDynix, to test improving visibility of library data on the web and user experiences by utilizing linked data (Ennis, 2016).