Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME

Karim Tharani

ABSTRACT

By way of a case study, this paper illustrates and evaluates the Bibliographic Framework (or BIBFRAME) as means for harvesting and sharing bibliographic metadata over the web for libraries. BIBFRAME is an emerging framework developed by the Library of Congress for bibliographic description based on Linked Data. Much like Semantic Web, the goal of Linked Data is to make the web "data aware" and transform the existing web of documents into a web of data. Linked Data leverages the existing web infrastructure and allows linking and sharing of structured data for human and machine consumption.

The BIBFRAME model attempts to contextualize the Linked Data technology for libraries. Library applications and systems contain high-quality structured metadata, but this data is generally static in its presentation and seldom integrated with other internal metadata sources or linked to external web resources. With BIBFRAME existing disparate library metadata sources such as catalogs and digital collections can be harvested and integrated over the web. In addition, bibliographic data enriched with Linked Data could offer richer navigational control and access points for users. With Linked Data principles, metadata from libraries could also become harvestable by search engines, transforming dormant catalogs and digital collections into active knowledge repositories. Thus experimenting with Linked Data using existing bibliographic metadata holds the potential to empower libraries to harness the reach of commercial search engines to continuously discover, navigate, and obtain new domain specific knowledge resources on the basis of their verified metadata.

The initial part of the paper introduces BIBFRAME and discusses Linked Data in the context of libraries. The final part of this paper outlines and illustrates a step-by-step process for implementing BIBFRAME with existing library metadata.

INTRODUCTION

Library applications and systems contain high-quality structured metadata, but this data is seldom integrated or linked with other web resources. This is adequately illustrated by the nominal presence of library metadata on the web.[1] Libraries have much to offer to the web and its evolving future. Making library metadata harvestable over the web may not only refine precision

Karim Tharani (karim.tharani@usask.ca) is Information Technology Librarian at the University of Saskatchewan in Saskatoon, Canada.

and recall but has the potential to empower libraries to harness the reach of commercial search engines to continuously discover, navigate, and obtain new domain specific knowledge resources on the basis of their verified metadata. This is a novel and feasible idea, but its implementation requires libraries to both step out of their comfort zones and to step up to the challenge of finding collaborative solutions to bridge the islands of information that we have created on the web for our users and ourselves.

By way of a case study, this paper illustrates and evaluates the Bibliographic Framework (or BIBFRAME) as means for harvesting and sharing bibliographic metadata over the web for libraries. BIBFRAME is an emerging framework developed under the auspices of the Library of Congress to exert bibliographic control over traditional and web resources in an increasingly digital world. While BIBFRAME has been introduced as a potential replacement for MARC (Machine-Readable Cataloging) in libraries;[2] however, the goal of this paper is to highlight the merits of BIBFRAME as a mechanism for libraries to share metadata over the web.

BIBFRAME and Linked Data

While the impetus behind BIBFRAME may have been replacement of MARC, "it seems likely that libraries will continue using MARC for years to come because that is what works with available library systems."[3] Despite its uncertain future in the cataloging world, BIBFRAME in its current form provides fresh and insightful mechanism for libraries to repackage and share bibliographic metadata over the web. BIBFRAME utilizes the Linked Data paradigm for publishing and sharing data over the web.[4] Much like Semantic Web, the goal of Linked Data is to make the web "data aware" and transform the existing web of documents into a web of data. Linked Data utilizes existing web infrastructure and allows linking and sharing of structured data for human and machine consumption. In a recent study to understand and reconcile various perspectives on the effectiveness of Linked Data, the authors raise intriguing questions about the possibilities of leveraging Linked Data for sharing library metadata over the web:

> *Although library metadata made the transition from card catalogs to online catalogs over 40 years ago, and although a primary source of information in today's world is the Web, metadata in our OPACs are no more free to interact on the Web today than when they were confined on 3" × 5" catalog cards in wooden drawers. What if we could set free the bound elements? That is, what if we could let serial titles, subjects, creators, dates, places, and other elements, interact independently with data on the Web to which they are related? What might be the possibilities of a statement-based, Linked Data environment?[5]*
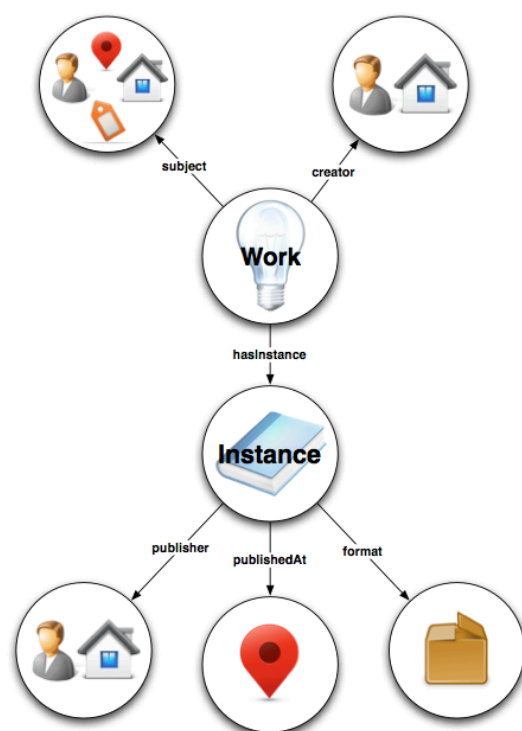
Figure 1. The BIBFRAME Model[6]

BIBFRAME provides the means for libraries to experiment with Linked Data to find answers to these questions for themselves. This makes BIBFRAME both daunting and delighting simultaneously. It is daunting because it imposes a paradigm shift in how libraries have historically managed, exchanged, and shared metadata. But embracing Linked Data also leads to a promise land where metadata within and among libraries can be exchanged seamlessly and economically over the web. BIBFRAME (http://bibframe.org) consists of a model and a vocabulary set specifically designed for bibliographic control.[7] The model identifies four main classes, namely, Work, Instance, Authority, and Annotation (see figure 1). For each of these classes, there are many hierarchical attributes that help in describing and linking instantiations of these classes. These properties are collectively called the BIBFRAME vocabulary.

Philosophically, Linked Data is based on the premise that more links among resources will lead to better contextualization and credibility of resources, which in turn will help in filtering irrelevant resources and discovering new and meaningful resources. At a more practical level, Linked Data provides a simple mechanism to make connections among pieces of information or resources over the web. More specifically, it not only allows humans to make use of these links but also machines to do so without human intervention. This may sound eerie, but one has to understand the history behind the origin of Linked Data not to think of this as yet another conspiracy for machines to take over the World (Wide Web).

In 1994 Tim Berners-Lee, the inventor of the web, put forth his vision of the Semantic Web as a "Web of actionable information—information derived from data through a Semantic theory for

interpreting the symbols. The Semantic theory provides an account of 'meaning' in which the logical connection of terms establishes interoperability between systems."[8] While the idea of Semantic Web has not been fully realized for a variety of functional and technical reasons, the notion of Linked Data introduced subsequently has made the concept much more accessible and feasible for a wider application.[9] Once again, it was Tim Berners-Lee who put forth the ground rules for publishing data on the web that are now known as the Linked Data Principles.[10] These principles advocate using standard mechanisms for naming each resource and their relationships with unique Universal Resource Identifiers (URIs); making use of the existing web infrastructure for connecting resources; and using Resource Description Framework (RDF) for documenting and sharing resources and their relationships.

A URI serves as a persistent name or handle for a resource and is ideally independent of the underlying location and technology of the resource. Although often used interchangeably, a URI is different from a URL (or Universal Resource Locator), which is a more commonly used term for web resources. A URL is a special type of URI, which points to the actual location (or the web address) of a resource, including the file name and extension (such as .html or .php) of a web resource. Being more generic, the use of URIs (as opposed to URLs) in Linked Data provides persistency and flexibility of not having to change the names and references every time resources are relocated or there is a change in server technology. For example if an organization switches its underlying web-scripting technology from Active Server Pages (ASP) to Java Server Pages (JSP), all the files on a web server will bear a different extension (e.g., .jsp) causing all previous URLs with old extension (e.g., .asp) to become invalid. This technology change, however, may have no impact if URIs are used instead of URLs because the underlying implementation and location details for a resource are masked from the public. Thus the URI naming scheme within an organization must be developed independent of the underlying technology. There are diverse best practices on how to name URIs to promote usability, longevity, and persistence.[11] The most important factors, however, remain the purpose and the context for which the resources are being harvested and shared.

Use of RDF is also a requirement of using Linked Data for sharing data over the web. Much like how HTML (Hypertext Markup Language) is used to create and publish documents over the web, RDF is used to create and publish Linked Data over the web. The format of RDF is very simple and makes use of three fundamental elements, namely, subject, predicate, and object. Similar to the structure of a basic sentence, the three elements make up the unit of description of a resource known as a *triple* in the RDF terminology. Unsurprisingly, RDF requires all three elements to be denoted by URIs with the exception of the object, which may also be represented by constant values such as a dates, strings, or numbers.[12] As an example, consider the work *Divine Comedy*. The fact this work, also known as *Divina Commedia*, was created by Dante Alighieri can be represented by the following two triples (using N-triples format):

```
<http://dbpedia.org/resource/Divine_Comedy>
<http://bibframe.org/vocab/creator>
<http://dbpedia.org/resource/Dante_Alighieri> .

<http://dbpedia.org/resource/Divine_Comedy>
<http://www.w3.org/2002/07/owl#sameAs>    "Divina Commedia".
```

In the first triple of this example, the work Divine Comedy (subject) is being attributed to a person called Dante Alighieria (object) as the creator (predicate). In the second triple the use of sameAs predicate asserts that both *Divine Comedy* and *Divina Commedia* refer to the same resource. Thus using URIs makes the resources and relationships persistent whereas use of RDF makes the format discernible by humans and machines. This seemingly simple idea allows data to be captured, formatted, shared, transmitted, received, and decoded over the web. Use of the existing web protocol (HTTP or Hypertext Transfer Protocol) for exchanging and integrating data saves the overhead of putting additional agreements and infrastructure in place among parties willing or wishing to exchange data. This ease and freedom to define relationships among resources over the web also makes it possible for disparate data sources to interact and integrate with each other openly and free of cost.

Why is this seemingly simple idea so significant for the future of the web? From a functional perspective, what this means is that Linked Data facilitates "using the Web to create *typed* links between data from different sources. These may be as diverse as databases maintained by two organisations in different geographical locations, or simply heterogeneous systems within one organisation that, historically, have not easily interoperated at the data level."[13] The notion of *typed linking* refers to the facility and freedom of being able to have and name multiple relationships among resources. From a technical point of view, "Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets."[14] In a traditional database, relationships between entities or resources are predefined by virtue of tables and column names. Moreover, data in such databases become part of the Deep Web and not readily accessed or indexed by search engines. [15]

The use of URIs to name relationships allows data sources to establish, use, and reuse vocabularies to define relationships between existing resources. These names or vocabularies, much like the resources they describe, have their own dedicated URIs, making it possible for resources to form long-term and reliable relationships with each other. If resources and relationships have and retain their identities by virtue of their URIs, then links between resources add to the awareness of these resources both for humans and machines. This is a key concept in realizing the overall mission of Linked Data to imbue data awareness and transforming the existing web of documents into a web of data. Consequently various institutions and industries

have established standard vocabularies and made them available for others to use with their data. For example, the Library of Congress has published its subject headings as Linked Data. The impetus behind this gesture is that if data from multiple organizations is "typed link" using LCSH (Library of Congress Subject Headings) with Linked Data, then libraries and others gain the ability to categorize, collocate, and integrate data from disparate systems over the Web by virtue of using a common vocabulary. As more and more resources link to each other through established and reusable vocabularies, the more data aware the Web becomes. Recognizing this opportunity, the Library of Congress has also developed and shared its vocabulary for bibliographic control as part of the BIBFRAME framework.[16]

Implementing BIBFRAME to Harvest and Share Bibliographic Metadata

Nowadays, systems like catalogs and digital collection repositories are commonplace in libraries, but these source systems often operate as islands of data both within and across libraries. The goal of this case study is to explore and evaluate BIBFRAME as a viable approach for libraries to integrate and share disparate metadata over the web. As discussed above, the BIBFRAME model attempts to contextualize the use of Linked Data for libraries and provides a conceptual model and underlying vocabulary to do so. To this end, a unique collection of Ismaili Muslim community was identified for the case study. The collection is physically housed at the Harvard University Library (HUL) and the metadata for the collection is dispersed across multiple systems within the library. An additional objective of this case study has been to define concrete and replicable steps for libraries to implement BIBFRAME. The discussion below is therefore presented in a step-by-step format for harvesting and sharing bibliographic metadata over the web.

1. *Establishing a Purpose for Harvesting Metadata*

The Harvard Collection of Ismaili Literature is first of its kind in North America. "The most important genre represented in the collection is that of the *ginans*, or the approximately one thousand hymn-like poems written in an assortment of Indian languages and dialects."[17] The feasibility of BIBFRAME was explored in this case study by creating a thematic research collection of ginans by harvesting existing bibliographic metadata at HUL. The purpose of this thematic research collection is to make ginans accessible to researchers and scholars for textual criticism. Historically libraries have played a vital role in making extant manuscripts and other primary sources accessible to scholars for textual criticism. The need for having such a collection in place for ginans was identified by Dr. Ali Asani, professor of Indo-Muslim and Islamic Religion and Cultures at Harvard University:

> *Perhaps the greatest obstacle for further studies on the ginan literature is the almost total absence of any kind of textual criticism on the literature. Thus far merely two out of the nearly one thousand compositions have been critically edited. Naturally, the availability of reliably edited texts is fundamental to any substantial scholarship in this field. . . . For the scholar of post-classical Ismaili literature, recourse to this kind of*

material has become especially critical with the growing awareness that there exist significant discrepancies between modern printed versions of several ginans and their original manuscript form. Fortunately, the Harvard collection is particularly strong in its holdings of a large number of first editions of printed ginan texts—a strength that should greatly facilitate comparisons between recensions of ginans and the preparation of critical editions.[18]

2. Modeling the Data to Fulfill Functional Requirements

Historically, the physicality of resources such as book or compact disc has dictated what is described in library catalogs and to what extent. The issue of cataloging serials and other works embedded within larger works has always been challenging for catalogers. For this case study as well, one of the major implementation decisions revolved around the granularity of defining a work. Designating each ginan as a work (rather than a manuscript or lithograph) was perhaps an unconventional decision, but one that was highly appropriate for the purpose of the collection. Thus there was a conscious and genuine effort to liberate a work from the confines of its carriers. Fortuitously, BIBFRAME does not shy away from this challenge and accommodates embedded and hierarchal works in its logical model. But BIBFRAME, like any other conceptual model, only provides a starting point, which needs to be adapted and implemented for individual project needs.
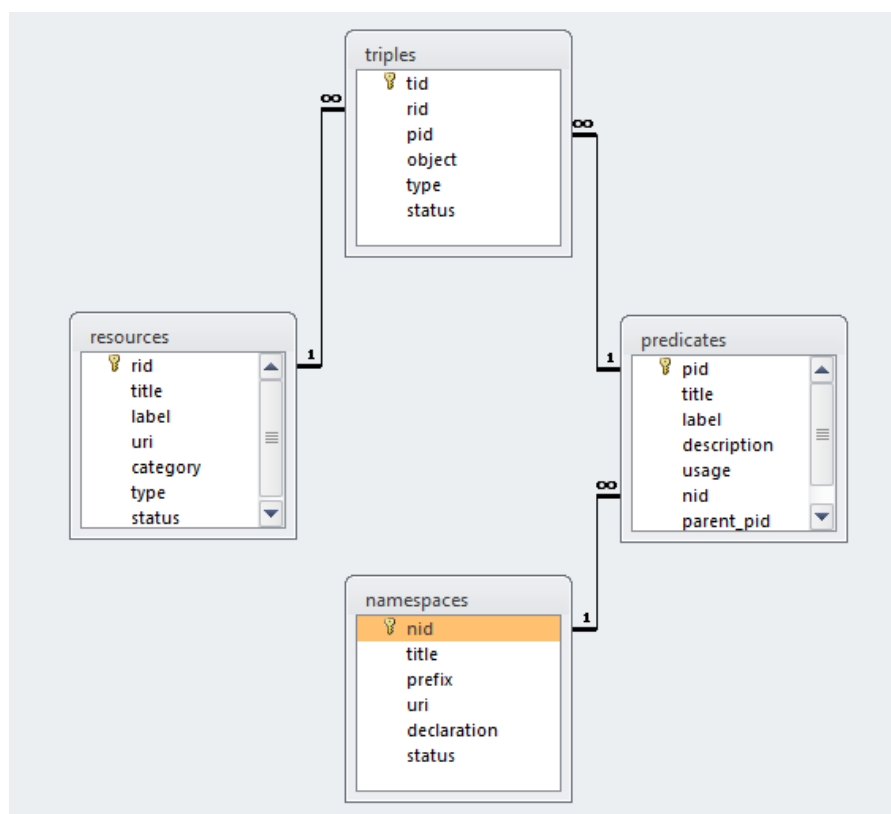
Figure 2. Excerpt of Project Data Model

The data model for this case study (see figure 2) was designed to balance the need to accommodate bibliographic metadata with the demands of Linked Data paradigm. Central to the project data model is the *resources* table where information on all resources along with their URIs and categories (work, instance, etc.) are stored. Resources relate to each other with use of *predicates* table, which captures relevant and applicable vocabularies. The *namespace* table keeps track of all the set of vocabularies being used for the project. In the *triples* table, resources are *typed linked* using appropriate predicates. Once the data model for the project was finalized, a database was created using MySQL to house the project data.

3. Planning the URI Scheme

In general the URI scheme for this case study conformed to the following intuitive nomenclature: <http://domain.com/resource/resource_type/resource_id>.

This URI naming scheme ensures that a URI assigned to a resource depends on its class and category (see table 1). While it may be customary to use textual identifiers in the URIs, the project used numeric identifiers to account for the fact that most of the ginans (works) are untitled and transliterated into English from various Indic languages. Generally support for using URIs is either already built-in or added on depending on the server technology being used. This case study utilized the LAMP (Linux, Apache, MySQL, and PHP) technology stack, and the URI handler for the project was added on to the Apache webserver using URL-rewriting (or mod_rewrite) facility.[19]

Resource Types	BIBFRAME Category	URI Example
Organizations	Annotation	http://domain.com/organization/1
Collections	Annotation	http://domain.com/collection/1
Items	Instance	http://domain.com/item/1
Ginan	Work	http://domain.com/ginan/1
Subjects	Authority	http://domain.com/subject/1

Table 1. URI Naming Scheme and Examples

4. Using Standard Vocabularies

BIBFRAME provides the relevant vocabulary and the underlying URIs to implement Linked Data with bibliographic data in libraries. While not all attributes may be applicable or used in a project, the ones that are identified as relevant must be referenced with their rightful URI. For example, the predicate hasAuthority from BIBFRAME has a persistent URI (http://bibframe.org/vocab/hasAuthority) enabling humans as well as machines to access and decode the purpose and scope of this predicate. Other vocabulary sets or namespaces commonly used with Linked Data include Resource Description Frameowrk (RDF), Web Ontology Language (OWL), Friend of a Friend (FOAF), etc. In rare circumstances, libraries may also choose to publish their own specific vocabulary. For example, any unique predicates for this case study could be

defined and published using the http://domain.com/vocab namespace.

5. Identifying Data Sources

The bibliographic metadata used for this case study was obtained from within HUL. As mentioned above, the data pertained to a unique collection of religious literature belonging to the Ismaili Muslim community of the Indian subcontinent. This collection was acquired by the Middle Eastern Department of the Harvard College Library in 1980. The collection comprises 28 manuscripts, 81 printed books, and 11 lithographs. In 1992, a book on the contents of this collection was published in 1992 by Dr. Asani and was titled *The Harvard Collection of Ismaili Literature in Indic Languages: A Descriptive Catalog and Finding Aid.* The indexes in the book served as one of the sources of data for this case study.

Subsequent to the publication of the book, the Harvard Collection of Ismaili Literature was also made available through Harvard's OPAC (online public access catalog) called HOLLIS (see figure 3). The catalog records were also obtained from the library for the case study. Some of the 120 items from the collection were subsequently digitized and shared as part of the Harvard's Islamic Heritage Project. The digital surrogates of these items were shared through the Harvard University Library Open Collections Program. and the library catalog records were also updated to provide

Figure 3. HOLLIS: Harvard University Library's OPAC

direct access to the digital copies where available. Additional metadata for the digitized items was also developed by the library to facilitate open digital access through Harvard Library's Page Delivery Service (PDS) to provide page-turning navigational interface for scanned page images over the web. Data from all these sources was leveraged for the case study.

6. *Transforming Source Metadata for Reuse*

ETL (Extract, Transform, and Load) is an acronym commonly used to refer to the steps needed to populate a target database by moving data from multiple and disparate source systems. Extraction is the process of getting the data out of the identified source systems and making it available for the exclusive use of the new database being designed. In the context of the library realm, this may mean getting MARC records out from a catalog or getting descriptive and administrative metadata out of a digital repository. Format in which data is extracted out of a source system is also an important aspect of the data extraction process. Use of XML (Extensible Markup Language) format is fairly common nowadays as most library source systems have built-in functionality to export data into a recognized XML standard such as MARCXML (MARC data encoded in XML), MODS (Metadata Object Description Schema), METS (Metadata Encoding and Transmission Standard), etc. In certain circumstances, data may be extracted using CSV (comma-separated values) format.

Transformation is the step in which data from one or more source systems is massaged and prepared to be loaded to a new database. The design of the new database often enforces new ways of organizing source data. The transformation process is responsible to make sure that the data from all source systems is integrated while retaining its integrity before being loaded to the new database. A simplistic example of data transformation may be that the new system may require authors' first and last names to be stored in separate fields rather than in a single field. How such transformations are automated will depend on the format of the source data as well as the infrastructure and programming skills available within an organization. Since XML is becoming the de facto standard for most data exchange, use of XSLT (Extensible Stylesheet Language Transformations) scripts is common. With XSLT, data in XML format can be manipulated and given different structure to aid in the transformation process.

The loading process is responsible for populating the newly minted database once all transformations have been applied. One of the major considerations in this process is maintaining the referential integrity of the data by observing the constraints dictated by the data model. This is achieved by making sure that records are correctly linked to each other and are loaded in proper sequence. For instance, to ensure referential integrity of items and their annotations, it may be necessary to load the items first and then the annotations with correct reference to the associated item identifiers.

For this case study, records from source systems were obtained in MARCXML and METS formats, and specific scripts were developed to extract desired elements and transform them into the required format. A somewhat unconventional mechanism was used to capture and reuse the data from Dr. Asani's book, which was only available in print. The entire book was scanned and processed by an OCR (Optical Character Recognition) tool to glean various data elements. Once the data was cleaned and verified, the information was transformed into a CSV data file to facilitate

database loading.

7. Generating RDF Triples

The RDF triples can be written or serialized using a variety of formats such as Turtle, N-Triples, JSON, as well as RDF/XML, among others. The traditional RDF/XML format, which was the first standard to be recommended for RDF serialization by the World Wide Web Consortium (W3C), was used for this case study (see figure 4). The format was chosen for its modularity in preserving the context of resources and their relationships as well as its readability for humans. Generating RDF may be a simple act if the data is already stored in a triplestore, which is a database specifically designed to store RDF data. But given that this project was implemented using a relational database management system (RDBMS), i.e., MySQL, the programming effort to generate RDF data was complex. The complications arose in identifying and tracking the hierarchical nature of the RDF data, especially in the chosen serialization format. Several server-side scripts were developed to aid in discerning the relationships among resources and formatting them to generate triples. In hindsight generating triples would have been easier using the N-triples serialization but that would have also required more complex programming for rebuilding the context for the user interface design.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bf="http://bibframe.org/vocab/" xmlns:gist="http://localhost/vocab/">
        <bf:Collection rdf:about="http://localhost/library/lod/resource/collection/1" type="collection">
                <bf:title>The Harvard Collection of Ismaili Literature in Indic Languages</bf:title>
                <bf:authorizedAccessPoint>Harvard Library Collection</bf:authorizedAccessPoint>
                <bf:label>Harvard Library Collection</bf:label>
                <bf:derivedFrom rdf:resource="http://hollis.harvard.edu/?itemid=|library/m/aleph|000602559"/>
                <bf:Organization rdf:resource="http://kat.usask.ca/library/lod/resource/organization/1"/>
                <bf:note type="description">The collection consists of 28 manuscripts, 81 letterpress volumes, and 11 lithographic volumes,
                        recording approximately 1,350 individual texts, primarily religious literature of the Shiite Ismaili community of the
                        Indian subcontinent....
                </bf:note>
        </bf:Collection>
        <bf:Organization rdf:about="http://klocalhost/library/lod/resource/organization/1" type="organization">
                <bf:title>Harvard University Library</bf:title>
                <bf:authorizedAccessPoint>Harvard Library</bf:authorizedAccessPoint>
                <bf:label>Harvard Library</bf:label>
                <bf:hasAuthority rdf:resource="http://id.loc.gov/vocabulary/organizations/mhh"/>
                <bf:derivedFrom rdf:resource="http://hcl.harvard.edu/libraries/houghton"/>
                <bf:note type="description">The collections of the Houghton Library focus on the study of Western civilization. Materials
                        relating to American, Continental, and English history and literature comprise the bulk of these collections and
                        include special concentrations ...
                </bf:note>
        </bf:Organization>
</rdf:RDF>
```

Figure 4. A Sample of Triples Serialized for the Project

8. Formatting RDF Triples for Human and Machine Consumption

The raw RDF data is sufficient for machines to parse and process, but humans typically require intuitive user interface to contextualize triples. In this case study, XSL was extensively used for formatting the triples. While XSLT and XSL (Extensible Stylesheet Language) are intricately related, they serve different purposes. XSLT is a scripting language to manipulate XML data whereas XSL is a formatting specification used in presentation of XML, much like how CSS (Cascading Style Sheets) are used for presenting HTML. A special routing script was also developed to detect whether the request for data was intended for machine or human consumption. For machine requests, the triples were served unformatted whereas for human requests, the triples were formatted to display in HTML.
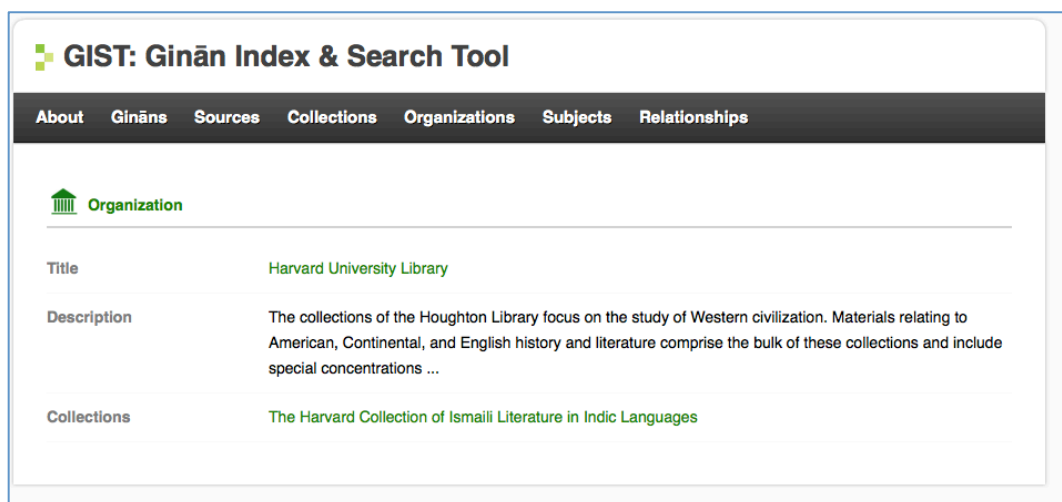
Figure 5. Formatted Triples for Human Consumption

DISCUSSION

Models are tools of communicating simple and complex relations between objects and entities of interest. Effectiveness of any model is often realized during implementation when the theoretical constructs of the models are put to test. The challenge faced by BIBFRAME, like any new model, is to establish its worthiness in the face of the existing legacy of MARC. The existing hold of MARC in libraries is so strong that it may take several years for BIBFRAME to be in a position to challenge the status quo. Historically bibliographic practices in libraries such as describing, classifying, and cataloging resources have primarily catered to tangible, print-based knowledge carriers such as books and journals.[20] BIBFRAME challenges libraries to revisit and refresh their traditional notion of text and textuality.

Although initially introduced as a replacement for MARC, BIBFRAME is far from being an either-or proposition given the MARC legacy. Nevertheless, BIBFRAME has made Linked Data paradigm much more accessible and practical for libraries. Rather than perceiving BIBFRAME as a threat to existing cataloging praxis, it may be useful for libraries to allow BIBFRAME to coexist within the current cataloging landscape as a means for sharing bibliographic data over the web. Libraries maintain and provide authentic metadata about knowledge resources for their users based on internationally recognized standards. This high quality structured metadata from library catalogs and other systems can be leveraged and repurposed to fulfill unmet and emerging needs of users. With Linked Data, library metadata could become readily harvestable by search engines, transforming dormant catalogs and collections into active knowledge repositories.

In this case study seemingly disparate library systems and data were integrated to provide a unified and enabling access to create a thematic research collection. It is also possible to create such purpose-specific digital libraries and collections as part of library operations without having to acquire additional hardware and commercial software. It was also evident from this case study that digital libraries built using BIBFRAME offer superior navigational control and access points

for users to actively interact with bibliographic data. Any Linked Data predicate has the potential to become an access point and act as a pivot to provide insightful view of the underlying bibliographic records (see figure 6). With advances in digital technologies "richer interaction is possible within the digital environment not only as more content is put within reach of the user, but also as more tools and services are put directly in the hands of the user."[21] Developing capacity to effectively respond to the informational needs of users is part and parcel of libraries' professional and operational responsibilities. With the ubiquity of the web and increased reliance of users on digital resources, libraries must constantly reevaluate and reimagine their services to remain responsive and relevant to their users.
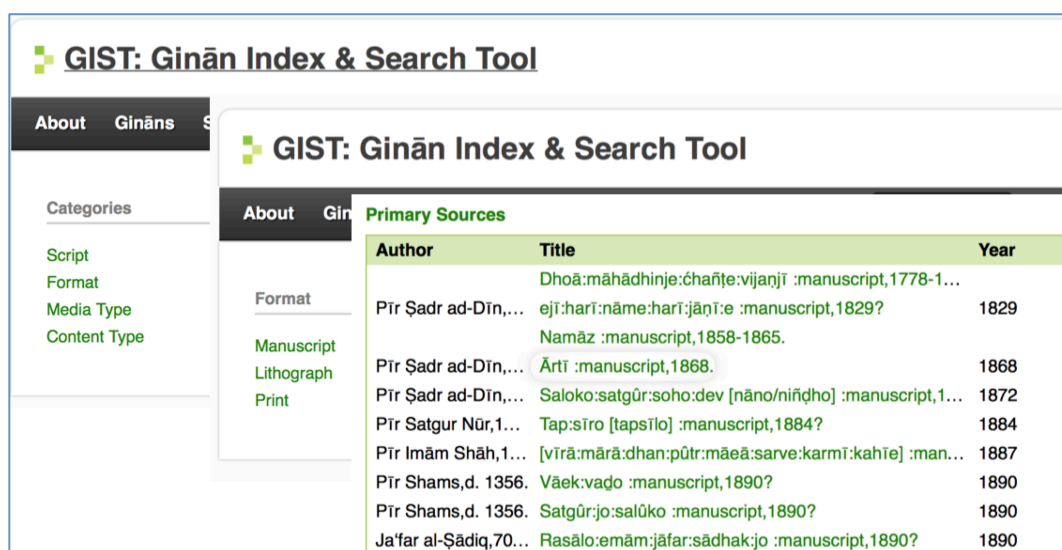

Figure 6. Increased Navigational Options with Linked Data

CONCLUSION

Just as libraries rely on vendors to develop, store, and share metadata for commercial books and journals, similar metadata partnerships need to be put in place across libraries. The benefits and implications of establishing such a collaborative metadata supply chain are far reaching and can also accommodate cultural and indigenous resources. Library digital collections typically showcase resources that are unique and rare, and the metadata to make these collections accessible must be shared over the web as part of library service.

As the amount of data on the web proliferates, users find it more and more difficult to differentiate between credible knowledge resources and other resources. BIBFRAME has the potential to address many of the issues that plague the web from a library and information science perspective, including precise search, authority control, classification, data portability, and disambiguation. Most popular search engines like Google are gearing up to automatically index and collocate disparate resources using Linked Data.[22] Libraries are particularly well positioned to realize this goal with their expertise in search, metadata generation, and ontology development. This research looks forward to further initiatives by libraries to become more responsive and make library

resources more relevant to the knowledge creation process.

REFERENCES

1. Tim F. Knight, "Break On Through to the Other Side: The Library and Linked Data," *TALL Quarterly* 30, no. 1 (2011): 1–7, http://hdl.handle.net/10315/6760.

2. Eric Miller et al., "Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services," November 11, 2012, http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf.

3. Angela Kroeger, "The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future," *Cataloging & Classification Quarterly* 51, no. 8 (2013): 873–89, http://dx.doi.org/10.1080/01639374.2013.823584.

4. Eric Miller et al., "Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services," November 11, 2012, http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf.

5. Nancy Fallgren et al., "The Missing Link: The Evolving Current State of Linked Data for Serials," *Serials Librarian* 66, no. 1–4 (2014): 123–38, http://dx.doi.org/10.1080/0361526X.2014.879690.

6. The figure has been adapted from Eric Miller et al., "Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services," November 11, 2012, http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf.

7. "Bibliographic Framework Initiative Project," Library of Congress, accessed August 15, 2014, http://www.loc.gov/bibframe.

8. Nigel Shadbolt, Wendy Hall, and Tim Berners-Lee, "The Semantic Web Revisited," *Intelligent Systems* 21 no. 3 (2006): 96–101, http://dx.doi.org/10.1109/MIS.2006.62.

9. Sören Auer et al., "Introduction to Linked Data and Its Lifecycle on the Web," in *Reasoning Web: Semantic Technologies for Intelligent Data Access*, edited by Sebastian Rudolph et al., 1–90 (Heidelberg: Springer, 2011), http://dx.doi.org/10.1007/978-3-642-23032-5_1.

10. Tim Berners-Lee, "Linked Data," *Design Issues*, last modified June 18, 2009, http://www.w3.org/DesignIssues/LinkedData.html.

11. Danny Ayers and Max Völkel, "Cool URIs for the Semantic Web," *World Wide Web Consortium (W3C),* last modified March 31, 2008, http://www.w3.org/TR/cooluris.

12. Tom Heath and Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space* (Morgan & Claypool, 2011), http://dx.doi.org/10.2200/S00334ED1V01Y201102WBE001.

13. Christian Bizer, Tom Heath, and Tim Berners-Lee, "Linked Data—The Story So Far," *International Journal on Semantic Web and Information Systems* 5, no. 3 (2009): 1–22, http://dx.doi.org/10.4018/jswis.2009081901.

14. Ibid.

15. Tony Boston, "Exposing the Deep Web to Increase Access to Library Collections" (paper presented at the AusWeb05, The Twelfth Australasian World Wide Web Conference, Queensland, Australia, 2005), http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1224/1509.

16. "Bibliographic Framework Initiative," BIBFRAME.ORG, accessed August 15, 2014, http://bibframe.org/vocab; "Bibliographic Framework Initiative Project," Library of Congress, accessed August 15, 2014, http://www.loc.gov/bibframe.

17. Ali Asani, *The Harvard Collection Ismaili Literature in Indic Languages: A Descriptive Catalog and Finding Aid* (Boston: G.K. Hall, 1992).

18. Ibid.

19. Ralf S. Engelschall, "URL Rewriting Guide," *Apache HTTP Server Documentation*, last modified December, 1997, http://httpd.apache.org/docs/2.0/misc/rewriteguide.html.

20. Yann Nicolas, "Folklore Requirements for Bibliographic Records: Oral Traditions and FRBR," *Cataloging & Classification Quarterly* 39, no. 3–4 (2005): 179–95, http://dx.doi.org/10.1300/J104v39n03_11.

21. Lee L. Zia, "Growing a National Learning Environments and Resources Network for Science, Mathematics, Engineering, and Technology Education: Current Issues and Opportunities for the NSDL Program," *D-Lib Magazine* 7, no. 3 (2001), http://www.dlib.org/dlib/march01/zia/03zia.html.

22. Thomas Steiner, Raphael Troncy, and Michael Hausenblas, "How Google is Using Linked Data Today and Vision for Tomorrow" (paper presented at the Linked Data in the Future Internet at the Future Internet Assembly (FIA 2010), Ghent, December 2010), http://research.google.com/pubs/pub37430.html.