Bibliographic Framework as a Web of Data:

Linked Data Model and Supporting Services

Library of Congress
Washington, DC
November 21, 2012

This document and model were developed under contract from the Library of Congress by a team from Zepheira composed of Eric Miller, Uche Ogbuji, Victoria Mueller, and Kathy MacDougall.

Table of Contents

Introduction

The Library of Congress officially launched its Bibliographic Framework Initiative in May 2011. The Initiative aims to re-envision and, in the long run, implement a new bibliographic environment for libraries that makes "the network" central and makes interconnectedness commonplace. Prompted in no small part by the desire to embrace new cataloging norms, it is essential that the library community redevelop its bibliographic data models as part of this Initiative. Toward that objective, this document presents a high-level model for the library community for evaluation and discussion, but it is also important to consider this document within a much broader context, and one that looks well beyond the library community.

Libraries have a long and rich history of using technology to realize economies of scale. The Library of Congress led the effort to develop and implement the Machine Readable Cataloging format (commonly known as MARC) in the 1960s to do just that - share cataloging information in electronic form. Instead of thousands of catalogers repeatedly describing the same resources, the effort of one cataloger could be shared with many. As time passed, the format remained largely unchanged as methods of exchange and data manipulation continued to evolve, thus that the library community has benefited from a substantial period of data model/format stability. It is a success story with few parallels. It leveraged new and yet-unproven technology, broke ground on standards development, led to widespread and lasting collaboration among libraries, and, crucially, reduced costs. It's time we do it again.

The new, proposed model is simply called BIBFRAME, short for Bibliographic Framework. The new model is more than a mere replacement for the library community's current model/format, MARC. It is the foundation for the future of bibliographic description that happens on, in, and as part of the web and the networked world we live in. It is designed to integrate with and engage in the wider information community while also serving the very specific needs of its maintenance community - libraries and similar memory organizations. It will realize these objectives in several ways:

1. Differentiate clearly between conceptual content and its physical manifestation(s) (e.g., works and instances)
2. Focus on unambiguously identifying information entities (e.g., authorities)
3. Leverage and expose relationships between and among entities

In a web-scale world, it is imperative to be able to cite library data in a way that not only differentiates the conceptual work (a title and author) from the physical details about that work's manifestation (page numbers, whether it has illustrations) but also clearly identifies entities involved in the creation of a resource (authors, publishers) and the concepts (subjects) associated with a resource. Standard library description practices, at least until now, have focused on creating catalog records that are independently understandable, by aggregating information about the conceptual work and its physical carrier and by relying heavily on the use of lexical strings for identifiers, such as the name of an author. The proposed BIBFRAME model encourages the creation of clearly identified entities and the use of machine-friendly identifiers which lend themselves to machine interpretation for those entities.

This is in keeping with contemporary, expected data practices, especially those integrating with the web. As the web is evolving from a network of linked documents to a network of linked documents and the data underpinning those documents, it is becoming clear that the data in those documents are crucial to helping the user locate information on the web. This very notion is the foundation for Schema.org, an effort led by Google, Bing, Yandex, and Yahoo! to create a universal vocabulary for web designers to better describe the information embedded in traditional web pages. When users begin their information hunt with a search engine or social network, whose objective is to help users locate information, then cultural heritage organizations need to help those engines and networks direct users to answers, especially those held by libraries. The BIBFRAME model is intentionally designed to coordinate the cataloging and metadata that libraries create with these efforts, and connect with them. In short, the BIBFRAME model is the library community's formal entry point for becoming part of a much larger web of data.

As libraries become part of this larger web of data, by leveraging the use of stable identifiers to reference clearly differentiated entities, focus will shift from capturing and recording descriptive details about library resources to identifying and establishing more relationships between and among resources. This includes related resources found on the web, and especially those beyond the traditional bounds of the library universe. These relationships - these links - drive the web, transforming the information space from many independent silos to a network graph that branches out in every direction. Relationships help search engines and other services to improve search relevancy and, most importantly, help users find the information they are looking for. Relationships are fundamental to the modern web as

exemplified by Google's Knowledge Graph which takes advantage of them to suggest to users other items they may be interested in based on their previous search, and by Amazon's relationships between buying patterns and book titles when calculating a list of additional items a customer may be interested in. Librarians are expert in identifying and capturing bibliographic relationships, and the BIBFRAME model establishes a basis to expose this expert knowledge. As our users search for relevant resources, bibliographic relationships will materially contribute to the richness of the web, and assist information seekers.

Although it is important to understand the Bibliographic Framework Initiative, and the proposed model, in a larger context, within the library community it is equally important to consider this document as a starting point upon which the community will continue to build. When reading about the BIBFRAME model, it is clear that much remains to be done. It is important to remember that this model, like MARC, must be able to accommodate any number of content models and specific implementations, but still enable data exchange between libraries. It needs to support new metadata rules and content standards that emerge, including the newest library content standard - RDA (Resource Description & Access). The BIBFRAME model must therefore both broaden and narrow the format universe for exchange of bibliographic data.

Although the BIBFRAME model is a draft and expected to change, the Library of Congress wants to share it now with the community not only so that it is informed of progress being made but also to engender conversation and constructive feedback. The Library is leading this initiative, but it is important that the library community as a whole work in concert to create an environment for bibliographic description and data exchange that recognizes and leverages the resources and scale of a global network of data. This needs to be achieved keeping in mind the resources and legacy data of libraries.

The Library looks forward to and welcomes feedback on the proposed new BIBFRAME model.

Library of Congress
November 19, 2012

Bibliographic Framework as a Linked Data Model

The Linked Data community in all its diversity draws inspiration from the thoughts of the Web's inventor, Tim Berners-Lee. In his article, "Giant Global Graph"[1] he expressed the basic evolution of the idea with a few apt observations:

The realization [behind creation of the Internet] was, "It isn't the cables, it is the computers which are interesting". The Net was designed to allow the computers to be seen without having to see the cables. The [World Wide Web] increases the power we have as users again. The realization was "It isn't the computers, but the documents which are interesting". Now you could browse around a sea of documents without having to worry about which computer they were stored on. Now, people are making another mental move. There is realization now, "It's not the documents, it is the things they are about which are important".

Berners-Lee points out that all stages of this evolution are about webs of links—a web of computers (though we call this a "network" rather than a "web"), a web of documents (what most people call "The Web"), and ultimately a web of all the things we want to share. He argues that we should extend the basic principles of the Web more directly to data (for example, the contents of traditional databases), and that we should not be shy about making links to non-computer resources such as people, tangible and abstract things, places, and so on. This expansive view of linking is called "the Web of data" and forms the basis of Linked Data.

The goal of this initial draft is to provide a pattern for modeling both future resources and bibliographic assets traditionally encoded in MARC 21. This pattern, which leverages the Web as an underlying architecture, will shape a common descriptive framework and achieve two objectives: (1) enable far more integration of existing bibliographic resources and (2) create a roadmap for moving forward toward refinement, redevelopment or development of alternative approaches. (There is more background on Linked Data and Linked Open Data on pp.22-27.)

---

[1] http://dig.csail.mit.edu/breadcrumbs/node/215

The Library community has been a pioneer in the creation, management, organization and collaborative curation of data related to the creative works our cultures have produced. MARC 21 is the latest evolution of a library interchange format that exchanges this data in a relatively specialized market. Libraries generate, maintain and curate an enormous amount of high-quality data, however, that is valuable well beyond traditional library boundaries. In reflecting the MARC 21 format to a Linked Data model we expand the utility and value of this data as well as the community that Libraries and Cultural Heritage institutions serve. In modeling the MARC 21 format as a Web of Data it is important to deconstruct and then reconstruct the informational assets that comprise MARC. MARC has a rich history in supporting the evolving needs of the library community that can be reflected in 3 primary functions[2]

1. Data related to the intellectual essence of a work
2. Data related to the actual instance of the work - that is what you hold in your hand, retrieve from an electronic source network, etc.
3. Record metadata such as control numbers, record handling codes and other annotations.

In deconstructing the data elements corresponding to these functions we can begin to materialize the concepts embedded in these data element sets as a **linkable information resource**. These "MARC Resources" can then be re-assembled into a coherent architecture that provides the basis of a new Bibliographic Framework and allows for cooperative cataloging at a more granular level (persons, places, subjects, organizations, etc.). Then, as we leverage the Web as an information architecture, whenever updates to these MARC Resources are performed (e.g. someone adds new information about a person, new mappings related to a subject, etc.) notification events can occur to automatically update systems that reference these Resources. Further, these information assets can now be more effectively utilized at a granular level and provide a richer substrate to which local collections, special collections and third party data can easily annotate and contextualize cooperative library content.

This document provides a draft Linked Data model for the new Bibliographic Framework and identifies a set of additional issues that will help shape the future of this work. This

---

[2] Delsey, T. (2002) "Functional Analysis of the MARC 21 Bibliographic and Holdings Format." www.loc.gov/marc/frbr/functional-analysis.html

Bibliographic Framework (BIBFRAME) Linked Data model is designed to be a basis for community discussion and dialog. It is not complete. Nor, if this model is effective, will it ever be "complete" as it can and should be extended to support a range of new applications and descriptive assets that we currently can not imagine.

The BIBFRAME Linked Data model attempts to balance the following factors:

- Flexibility to accommodate future cataloguing domains, and entirely new use scenarios and sources of information
- The Web as an architectural model for expressing and connecting decentralized information
- Social and technical adoption outside the Library community
- Social and technical deployment within the Library community
- Previous efforts in expressing bibliographic material as Linked Data
- Application of machine technology for mechanical tasks while amply accommodating the subject matter expert (the librarian) as the explicit brain behind the mechanics.
- Previous efforts for modeling bibliographic information in the library, publishing, archival and museum communities
- The robust and beneficial history and aspects of a common method of bibliographic information transfer

The BIBFRAME model

The BIBFRAME Model consists of the following main classes:

- **Creative Work** - a resource reflecting a conceptual essence of the cataloging item.
- **Instance** - a resource reflecting an individual, material embodiment of the Work.
- **Authority** - a resource reflecting key authority concepts that have defined relationships reflected in the Work and Instance. Examples of Authority Resources include People, Places, Topics, Organizations, etc.
- **Annotation** - a resource that decorates other BIBFRAME resources with additional information. Examples of such annotations include Library Holdings information, cover art and reviews.

A graphical representation of this high level model and relationships among these core BIBFRAME classes is illustrated in Figure 1.

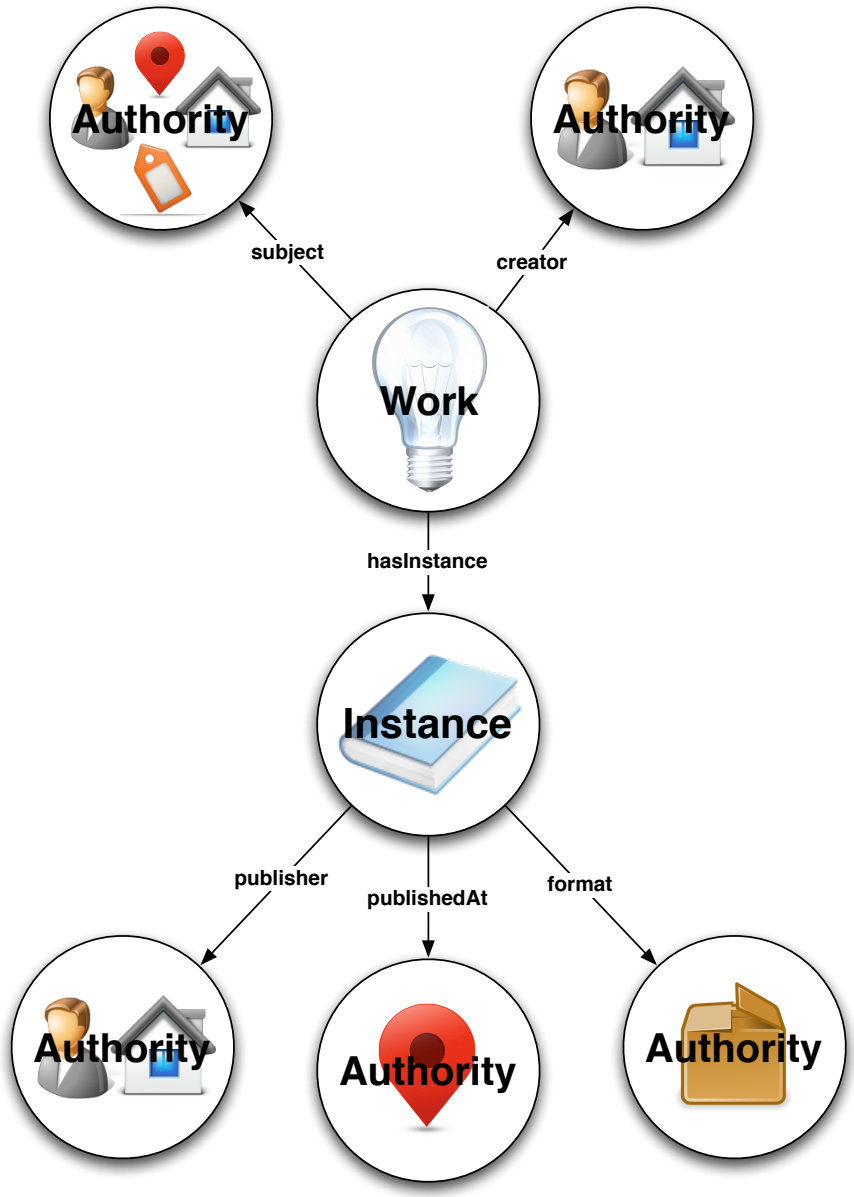

Figure 1: A graphical representation of the BIBFRAME Linked Data model defining the relation between Work and Instance resources and their contextualization to Web addressable Authority resources.

In the context of entity relationship models, including FRBR, the BIBFRAME model recognizes entities, attributes, and relationships between entities. BIBFRAME leverages the World Wide Web Consortium's Resource Description Framework (RDF) modeling practice of uniquely identifying as Web resources all entities (resources), attributes, and relationships between entities (properties). This allows for further annotations (such as mappings to other vocabularies or local community extensions) to be enabled as needed.

BIBFRAME Creative Works

A BIBFRAME Creative Work, abbreviated simply as Work, reflects a conceptual cataloging item. A Work is an abstract entity as there is no single material object one can point to. The Work exists as a Web based control point that reflects both commonality of content between and among the various Instances associated with the Work as well as a reference point for other Works. Common properties of Works include contextual relationships to BIBFRAME Authorities related to the "subjectness" (Topic, Person, Place, Geographical, etc.) of the resource as well as the entities (Person, Organization, Meeting, etc.) associated with its creation. Works can relate to other Works reflecting, for example, part / whole relationships.

BIBFRAME Instances

BIBFRAME Instances reflect an individual, material embodiment of a BIBFRAME Work that can be physical or digital in nature. A BIBFRAME Instance exists as a Web based control point that includes properties specific to the materialization as well as contextual relationships to appropriate BIBFRAME Authorities related to the publication, production, distribution of the material resource. Each BIBFRAME Instance is an instance of one and only one BIBFRAME Work.

BIBFRAME Authorities

BIBFRAME Authorities are key authority concepts that are the target of defined relationships reflected in the Work and Instance. Example of BIBFRAME Authority Resources include People, Places, Topics, Organizations, etc. From a cataloging perspective Authorities provide a means for supporting disambiguation and synchronization around authoritative information. From a users perspective, BIBFRAME Authorities provide effective and efficient control points that can be used to help navigate and contextualize related BIBFRAME Works and Instances. BIBFRAME

"BIBFRAME Authorities are not designed to compete or replace existing authority efforts but rather provide a common abstraction layer over various different Web based authority efforts to make them even more effective."

Authorities are not designed to compete or replace existing authority efforts but rather provide a common, light weight abstraction layer over various different Web based authority efforts to make them even more effective.

Annotating the BIBFRAME model

Libraries generate, maintain and improve an enormous amount of high-quality data that is valuable well beyond traditional library boundaries. The Bibliographic Framework Initiative recognizes this by including as a goal the ability to "accommodate and distinguish expert-, automated-, and self-generated metadata, including annotations (reviews, comments) and usage data."[3] Rather than pre-define and limit our potential uses of this data, the BIBFRAME model provides the necessary scaffolding to allows this data to easily be annotated by libraries as well as third party users of this information.

Figure 2 is a graphical representation of the BIBFRAME Linked Data model in the context of a flexible annotation framework. In this example, "holdings information" (who holds an instance and where such an instance might be found) is not a characteristic of the instance itself, but rather an assertion that a particular library makes about such an instance. Reflecting this as an annotation allows libraries to assert more localized descriptive metadata about the instance including reorganization, localized patron annotations, usage data, access policies, etc. Credible 'reviews', which in Figure 2 is reflected as an annotation of a Work, is another simple yet powerful example of this annotation model. These kind of annotations can be made "in-band" (for example by a particular Library adding their own local review of a work) or "out of band" (in terms of subscribing to and overlaying the data from a 3rd party review service). In any case Linked Data principles make such annotations easy to extend, represent and manage.

"Reflecting holdings as an annotation allows libraries to assert more localized descriptive metadata about the instance."

The ability for institutions to subscribe to cover art services, overlay their own versions of cover art or allow authorized users to contribute images are further, practical examples that are enabled by such an annotation model. The combination of these three kinds of

---

[3] A Bibliographic Framework for the Digital Age (October 31, 2011) http://www.loc.gov/marc/transition/news/framework-103111.html

examples is currently found on such commercial services as Amazon and can easily be adapted to support a more useful user experience for library and cultural heritage patrons. Figure 3 demonstrates using annotations on BIBFRAME instances to merge BIBFRAME data that describes an object in a library collection and an authorized user contributed image.

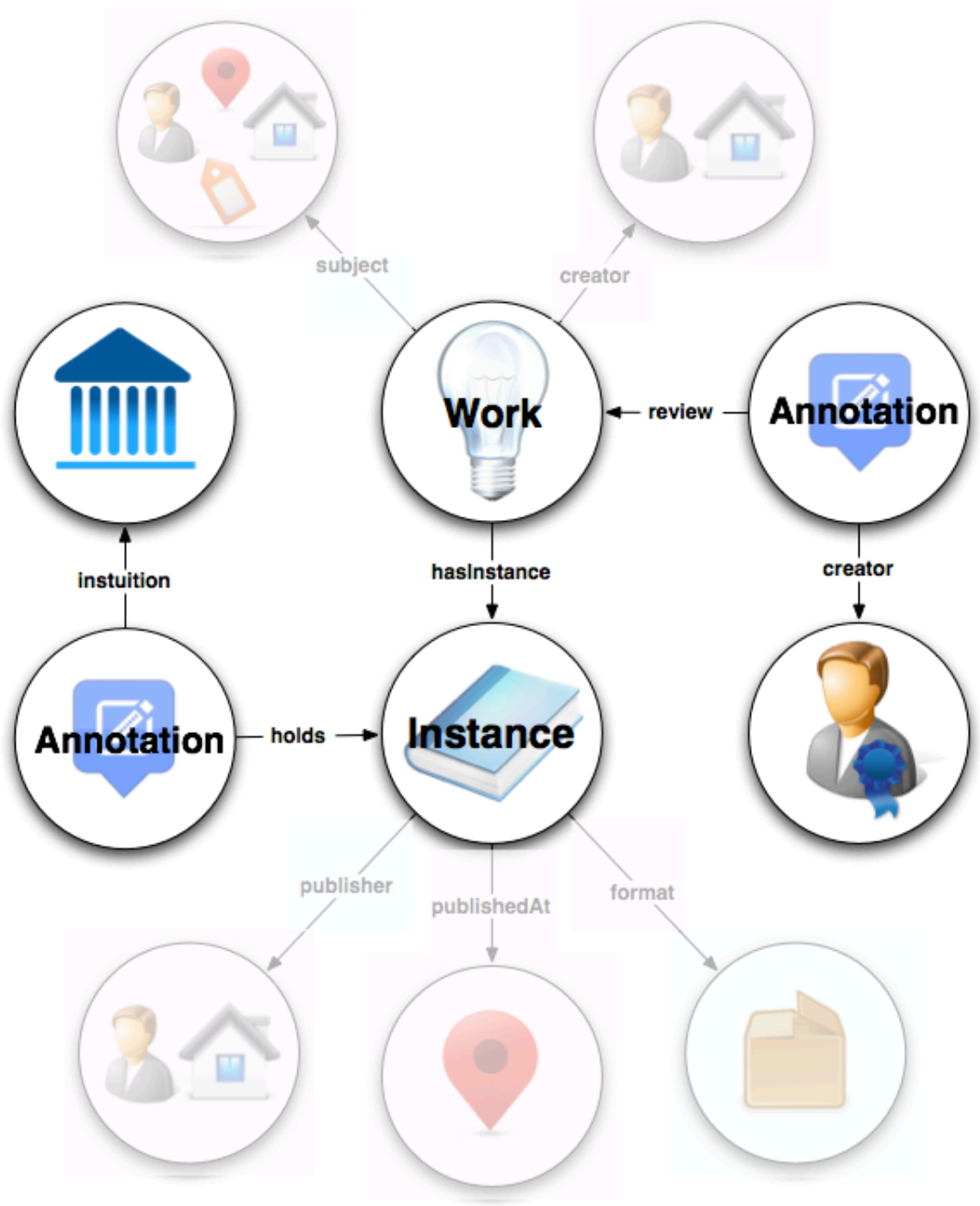

Figure 2: A graphical representation of the BIBFRAME Linked Data model in the context of a flexible annotation framework.

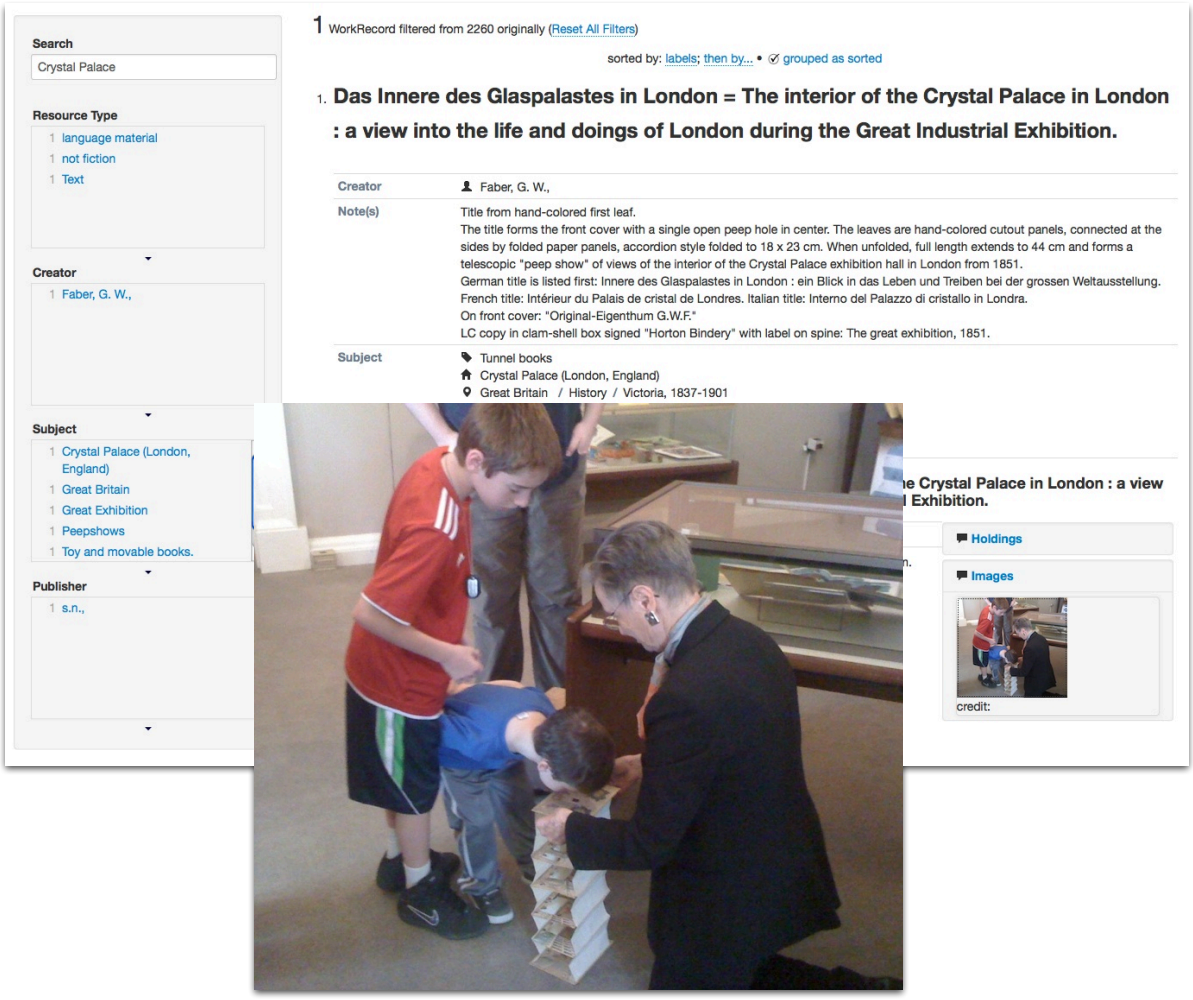

Figure 3: User contributed photograph as annotation of BIBFRAME instances.

Another practical example of this annotation model can be found in the increasing desire to find credible information in the digital age. Projects such as Reference Extract[4] are designed to capture references that librarians and other experts use in answering questions in such services as Library of Congress and OCLC's Question Point service [5]. This information, including data used to determine the most credible resources, is harvested, processed and then made available through a variety of Web environments. Figure 4 shows a snapshot of a

[4] MacArthur Foundation funds 'Reference Extract' to draw on librarians' expertise and add credibility to Web search experience http://www.oclc.org/news/releases/2011/20111.htm

[5] Question Point http://www.questionpoint.org/

prototype developed in the Reference Extract project which overlays the results of a traditional search engine with these credibility annotations. The ability to easily annotate library resources as credible answers to patron driven questions allows for new services to be developed that overlay these answers in traditional search engine searches to help end users determine resources that are credible.

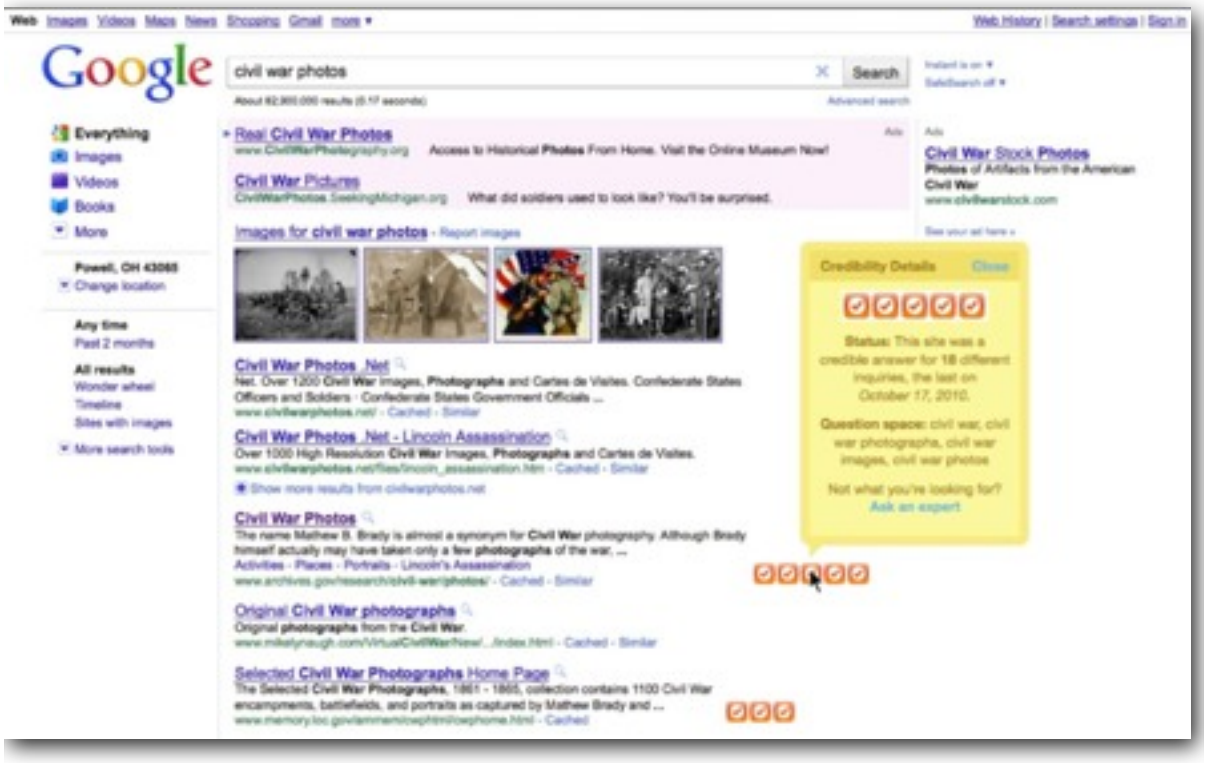

Figure 4: A credibility overlay of a search on 'civil war photos'.

Some other initiatives into next-generation cataloguing have picked up on basics of Linked Data, but have not fully worked through the architecture for expressing and aggregating such layered annotations. One example of where this leads to a practical problem is in workflow, and in particular the copying and sharing of records. BIBFRAME is unique in providing a built-in framework for annotating records with such lifecycle events, which allows one to deal with problems such as tracing errors to a source, or determining the actual holdings information associated with an instance. This is similar to what's called provenance in Museums and Archiving, and even in some technical Semantic Web contexts, but the BIBFRAME approach is a simpler one of annotating workflow as the record is used.

The capability of compounding resources with annotations and eventually services is a hallmark of the Web and how it has revolutionized many areas of inquiry. The annotation approach of BIBFRAME is key to creating a system driven by the knowledge capacity of libraries rather than the commercial interests that presently dominate the Web.

"The annotation approach of BIBFRAME is key to creating a system driven by the knowledge capacity of libraries rather than the commercial interests that presently dominate the Web."

The goal of the Bibliographic Framework Initiative is to develop a model to which various content models can be mapped. This recognizes that different communities may have different views of their resources and thus different needs for resource descriptions. This is especially pronounced as one leaves the book/text media and considers images (still and moving), cartographic resources, archival collections, and ultimately cultural artifact and museum collections. Many content models define hierarchical relationships that need to be restated in RDF graph terms and then simplified to the BIBFRAME model.

For example, the origin of the Work/Instance aspects of the BIBFRAME can reflect the FRBR relationships in terms of a graph rather than as hierarchical relationships, after applying a reductionist technique to simplify things as much as possible. Formally reconciling the BIBFRAME modeling effort with an RDA-lite set of cataloging rules is a logical next step.

The BIBFRAME Vocabulary

The BIBFRAME model is defined in RDF. The RDF vocabulary for the draft BIBFRAME model, to be available shortly, will provide a convenient way of navigating around the RDF model.

The suggested namespace for the BIBFRAME model is http://loc.gov/bibframe/vocab. An RDF/XML serialization of this vocabulary is expected to be available at this location. While the recommendation of a singular namespace is counter to several current Linked Data bibliographic efforts, it is crucial to clarify responsibility and authority behind the schematic framework of BIBFRAME in order to minimize confusion and reduce the complexity of the resulting data formats. It will be the role of the Library's standards stakeholders to maintain the connections between BIBFRAME model elements and source vocabularies such as

Dublin Core, FOAF, SKOS and future, related vocabularies that may be developed to support different aspects of the Library workflow.

Identifier persistence is not a technology issue but rather an organizational, policy, and community one. A persistence policy should be defined stating clearly the persistence and change management mechanisms.

Serializing the BIBFRAME model

There can be several serializations of the BIBFRAME Linked Data model. The following XML serialization (of the RDF data model), while subject to change, is provided as a concrete example. This example is designed to provide a serialized encoding of a particular Work, its corresponding Instances and associated Authority information. The Work in question is the 'Functional Requirements for Bibliographic Records: Final Report'. The original BIBFRAME record associated with item is available here http://lccn.loc.gov/2001433363. Three Instances (one physical, one PDF, and one HTML web site) along with the associated Authority information (subjects, authors, publishers, etc.) are included in this example. The following example does not reflect a full MARC 21 to BIBFRAME mapping. Links in the following examples are included to illustrate the use of using URLs for defining BIBFRAME resources, the URLs themselves are not valid.

```
<!-- Work -->
<Report id = "http://bibframe/work/frbr-report">
  <title>Functional requirements for bibliographic records :</title>
  <titleRemainder>final report / IFLA Study Group on the Functional
     Requirements for Bibliographic Records ; approved by the Standing
     Committee of the IFLA Section on Cataloguing.</titleRemainder>

  <creator resource = "http://bibframe/auth/org/ifla" />
  <subject resource = "http://bibframe/auth/topic/cataloging" />
  <subject resource = "http://bibframe/auth/topic/bibliography" />
  <subject resource = "http://bibframe/auth/topic/frbr" />
  <abstract>The purpose of this study is to delineate in clearly defined
     terms the functions performed by the bibliographic record with respect
     to various media, various applications, and various user needs. The
     study is to cover the full range of functions for the bibliographic
     record in its widest sense- i.e., a record that encompasses not only
     descriptive elements, but access points (name, title, subject, etc.),
```

```
    other 'organizing' elements (classification, etc.), and annotations.
    </abstract>

  <language>English</language>

  <hasInstance resource="http://bibframe/inst/frbr-1997-09-01:0" />
  <hasInstance resource="http://bibframe/inst/frbr-1997-09-01:1" />
  <hasInstance resource="http://bibframe/inst/frbr-1997-09-01:2" />
</Report>


<!-- Instance -->
<HardcoverBook id="http://bibframe/inst/frbr-1997-09-01:0">
  <date>1998</date>
  <place resource=”http://bibframe/auth/geo/münchen” />
  <publisher resource="http://bibframe/auth/org/k.g.saur" />
  <isbn>359811382X</isbn>
</HardcoverBook>

<!-- Instance -->
<DigitalResource id="http://bibframe/inst/frbr-1997-09-01:1">
  <link>http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf</link>
  <format>application/pdf</format>
  <date>1997-09-01</date>
  <publisher resource="http://bibframe/auth/org/ifla" />
</DigitalResource>


<!-- Instance -->
<DigitalResource id="http://bibframe/inst/frbr-1997-09-01:2">
  <link>http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm</link>
  <format>text/html</format>
  <date>2007-12-26</date>
  <publisher resource="http://bibframe/auth/org/ifla" />
</DigitalResource>

<!-- BIBFRAME Topic -->
<Topic id="http://bibframe/auth/topic/frbr">
  <label>FRBR (Conceptual model)</label>
```

```
  <hasIDLink resource="http://id.loc.gov/authorities/subjects/
      sh2007002541" />

</Topic>


<!-- BIBFRAME Topic -->

<Topic id="http://bibframe/auth/topic/bibliography">

  <label>Bibliography</label>

  <generalSubdivision>Methodology</generalSubdivision>

  <hasIDLink resource=”http://id.loc.gov/authorities/subjects/
      sh85013838” />

</Topic>


<!-- BIBFRAME Topic -->

<Topic id="http://bibframe/auth/topic/cataloging">

  <label>Cataloging</label>

  <hasIDLink resource=”http://id.loc.gov/authorities/subjects/
      sh85020816” />

</Topic>


<!-- BIBFRAME Organization -->

<Organization id="http://bibframe/auth/org/ifla">

  <label>IFLA Study Group on the Functional Requirements for Bibliographic
     Records</label>

  <link>http://www.ifla.org/</link>

  <hasIDLink resource="http://id.loc.gov/authorities/names/nr98013265” />

</Organization>


<!-- BIBFRAME Organization -->

<Organization id="http://bibframe/auth/org/k.g.saur">

  <label>K.G. Saur</label>

  <link>http://www.degruyter.com/</link>

  <hasIDLink resource="http://id.loc.gov/authorities/names/nr91037301” />

</Organization>


<!-- BIBFRAME Place -->
```

```
<Place id="http://bibframe/auth/geo/münchen">
  <label>Munich (Germany)</label>
  <hasIDLink resource="http://id.loc.gov/authorities/names/n79059670" />
</Place>
```

A high level RDF model reflecting the relationship between the Work and the corresponding Instances as defined by this XML serialization is shown in Figure 5.

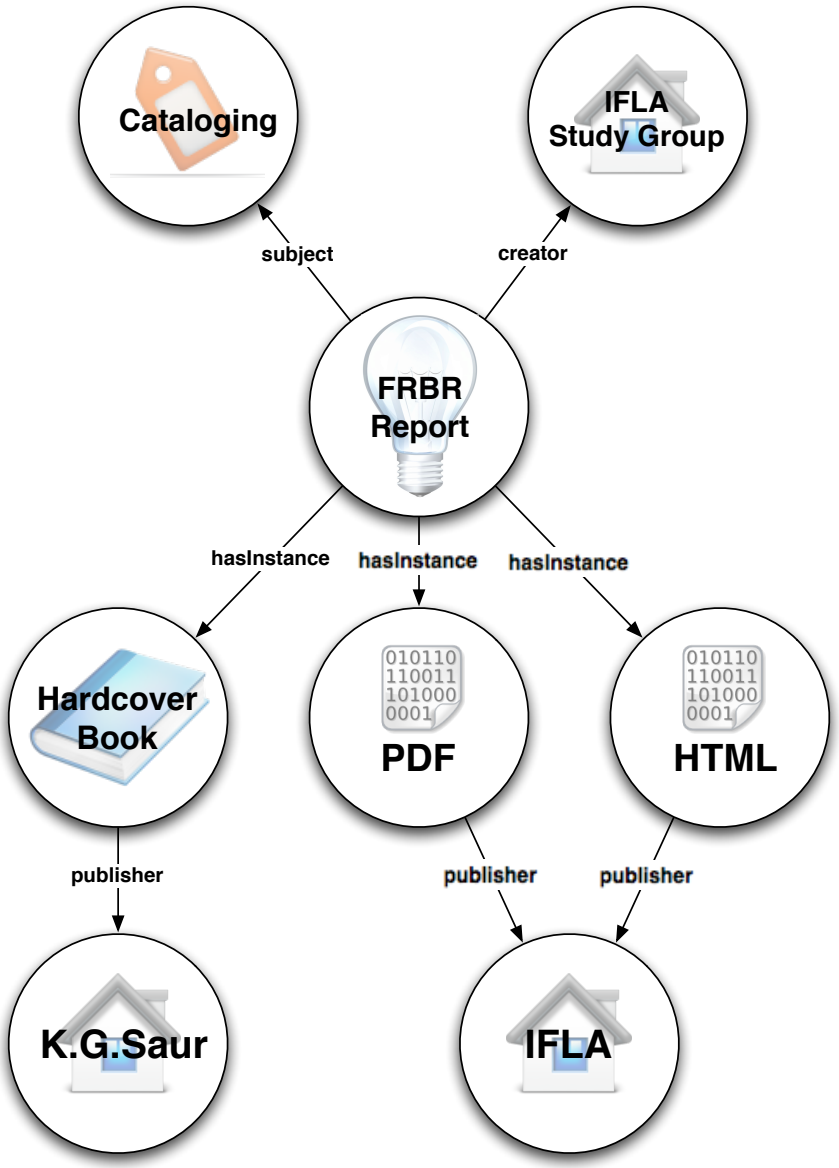

Figure 5: A high level RDF model associated with the XML serialization. The FRBR Report has 3 instances that are associated with 2 different publishers.

Given a sample serialized encoding of this Work, its corresponding Instances and associated Authority information, the following example serialization represents a particular library holdings annotation for the 'Functional Requirements for Bibliographic Records: Final Report' book published by K.G. Saur.

```
<!-- Holdings Annotation -->
<Holdings id="http://bibframe/annot/holdings/frbr-1997-09-01:0">
  <annotates resource="http://bibframe/inst/frbr-1997-09-01:0" />
  <institution resource=" http://bibframe/auth/org/
     ohio.university.alden" />
  <callNumber>025.3 F979 1998</callNumber>
  <access>circulating</access>
  <status>available</status>
</Holdings>

<!-- BIBFRAME Organization -->
<Organization id="http://bibframe/auth/org/ohio.university.alden">
  <label>Ohio University, Alden Library</label>
  <city>Athens</city>
  <state>OH</state>
  <zip>45701</zip>
  <link>http://www.library.ohiou.edu/</link>
  <hasIDLink resource="http://id.loc.gov/authorities/names/n2003039990" />
</Organization>
```

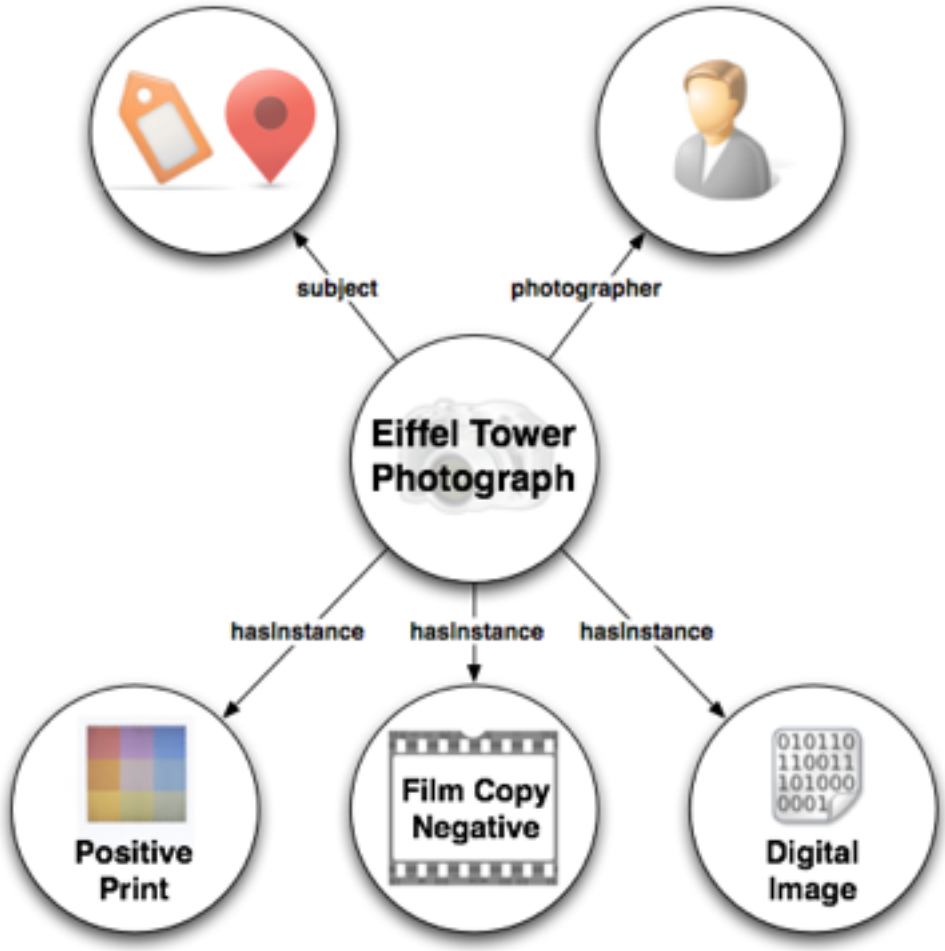

Figure 6: Example of a non-textual resource. The photograph has 3 instances associated with it.

Supporting Tools

A key part of supporting the BIBFRAME model is in providing tools and supporting services for helping migrate from MARC to a Linked Data environment. They should provide a means of navigating the output of a declarative BIBFRAME pipeline which takes existing MARC 21 data and translates this to the BIBFRAME model. Figure 7 for example provides a sample faceted Linked Data interface over this BIBFRAME data using Exhibit[6] which is a lightweight, open source, publishing framework for data-rich Interactive Web Pages tool that was developed by the MIT Simile Project.

[6] Simile Exhibit - http://www.simile-widgets.org/exhibit/

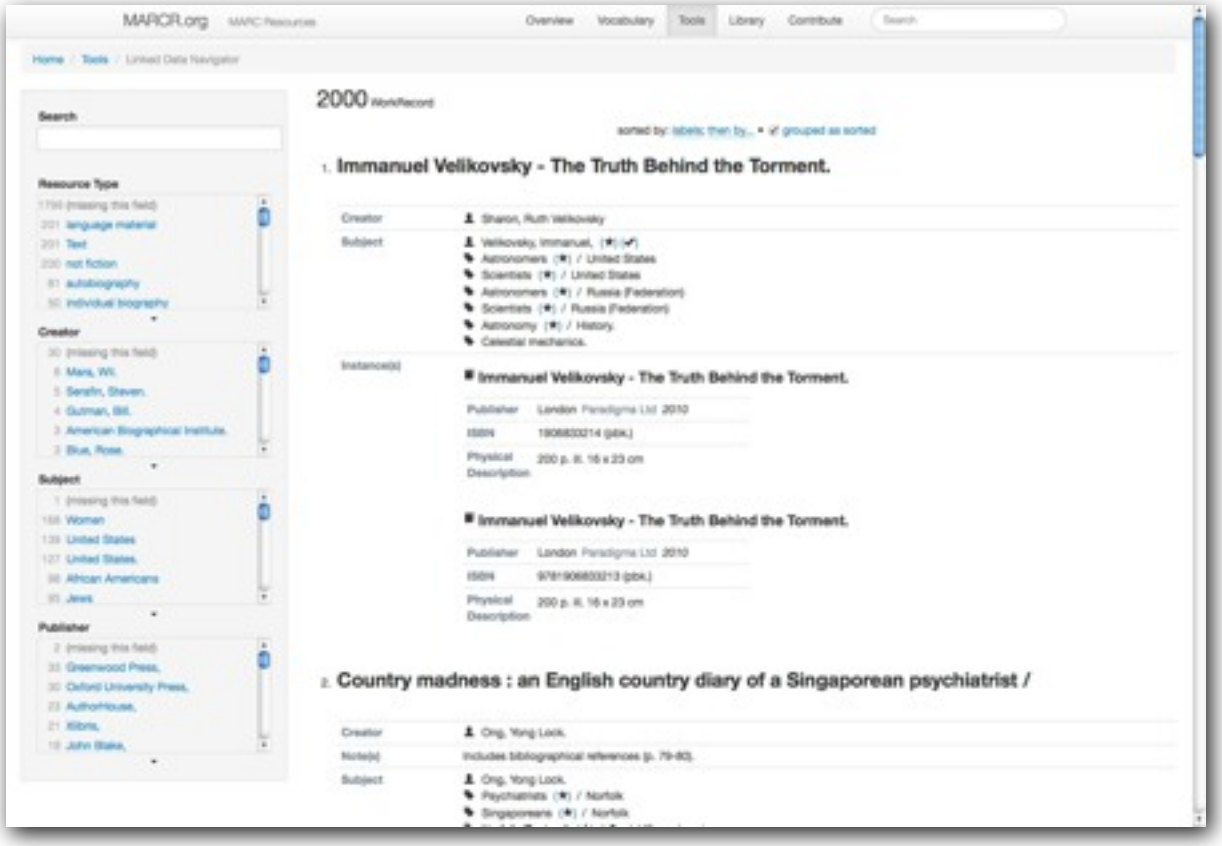

Figure 7: A faceted Linked Data interface over MARC 21 data reflected in the BIBFRAME model.

Figure 8 illustrates an expanded view of the use of BIBFRAME Authorities for the creator and subject of the Work. Visual cues (stars and check marks) reflect linkages that have been established from the MARC resources materialized from the original MARC 21 records and Linkable authority files. Linkable authority services that are currently used in this pipeline include the Library of Congress Linked Data Service (ID)[7] and VIAF Virtual International Authority File[8].

[7] Library of Congress Linked Data Service (ID) - http://id.loc.gov/

[8] VIAF Virtual International Authority File - http://viaf.org/

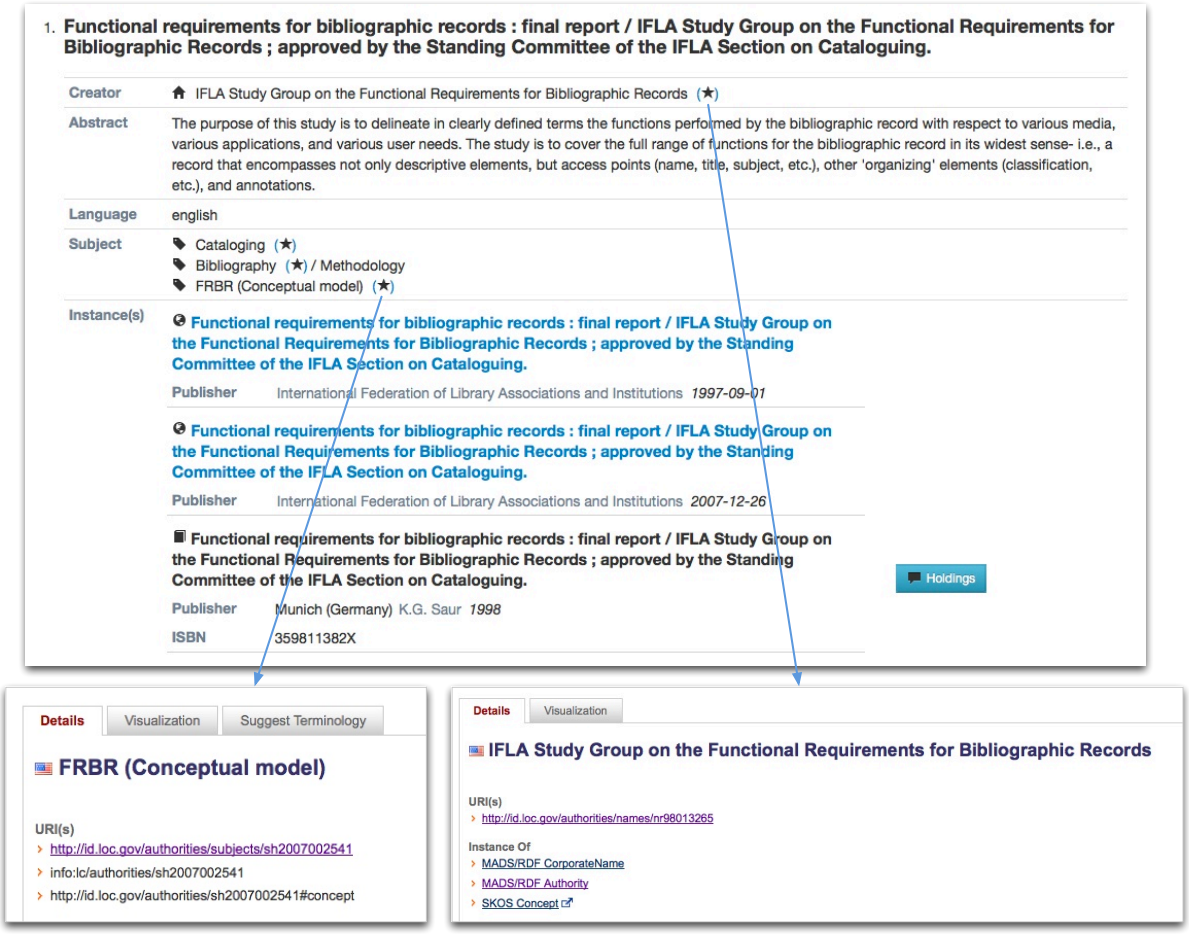

Figure 8: An expanded view of the Creator and Subject values. The stars reflect linkages that have been established from the MARC resources materialized from the original MARC 21 records and Linkable authority files at http://id.loc.gov/.

Background on Linked Data and LOD

An informal community, Linking Open Data (LOD), sprang up around the principles for a practical Data Web outlined by Tim Berners-Lee in a draft document on Linked Data[9]. The W3C has since provided support for this community, combining the vision of the W3C for

[9] http://www.w3.org/DesignIssues/LinkedData.html

using semantic features to enhance the Web with the pragmatism that characterizes mainstream Web 2.0. As the W3C-hosted LOD wiki says:

The goal of the W3C SWEO [Semantic Web Education and Outreach] Linking Open Data community project is to extend the Web with a data commons by publishing various open datasets as RDF on the Web and by setting RDF links between data items from different data sources.

The emphasis on RDF is natural for the W3C, which has been advocating the technology for a decade, but one development that gives LOD extra legs is the emergence of influential voices realizing that insistence on strict RDF format across the board is probably not the best present strategy for winning over Web developers. LOD supports RDF as a conceptual model, but the emphasis is more on inter-linking and shared patterns than on any one syntax. The full LOD community is a penumbra around the W3C-led core who support all the advantages of opening up data, and who see RDF, Atom, JSON, and so on as merely tools for Web developers to enrich their data.

Linked Open Data (LOD) 101

Taking a closer look at what LOD means in practice, the starting point is in four basic principles Berners-Lee drafted in his, "Linked Data" paper. Paraphrased, these are:

- Use URIs to identify things that you expose to the Web as resources.
- Use HTTP URIs so that people can locate and look up (dereference) these things.
- Provide useful information about the resource when its URI is dereferenced.
- Include links to other, related URIs in the exposed data as a means of improving information discovery on the Web.

Principle 1 means you should try to expose information as much as possible using URIs. This should be done not only for Web pages, but for front-office application documents, database rows and metadata, personal data, transactional logs, business rules and policies, and even services. If sharing it is useful, look to give its component pieces URIs. With regard to security, perhaps you are used to relying on the traditional application to protect your data. Remember that people bank on the Web. They make stock trades on the Web.

They book travel and buy things on the Web. The Web is well-proven as a secure data conduit, provided best practices are followed.

Principle 2 means you should give up obscure ID schemes (even URI schemes) and just stick to HTTP, which has served the Web so well. This ensures that the widest variety of tools and resources will be able to access it.

Principle 3 means that the data that you provide to people who access the data URIs should be in a common format suitable for sharing on the Web. XML is one obvious candidate for this, but not all XML is suitable. You need to use XML in a way that is semantically transparent, which means that the constructs in the XML are described in a rich way that can be processed by a machine. RDF is the main format used in the LOD community. It offers very high semantic transparency, but support for RDF is not yet as widespread as support for XML. One way to get the best of both worlds is to use GRDDL, a system for viewing XML through an RDF lens.

Principle 4 is the "share the wealth" principle. The first three encourage you to make possible Web pointers to data, and to maximize the usefulness of the data at those pointers. Once you have these pointers, you should not be shy about using them. Provide links as broadly as you can. You never know how someone or some machine is going to choose to navigate your Web of data, and the entire goal of LOD is to make it easy to use data in ways that were not originally conceived.

Expanding bubbles

The LOD community maintains a diagram of significant, public data sets that are available using LOD principles. Figure 9 is a recent version of that diagram.

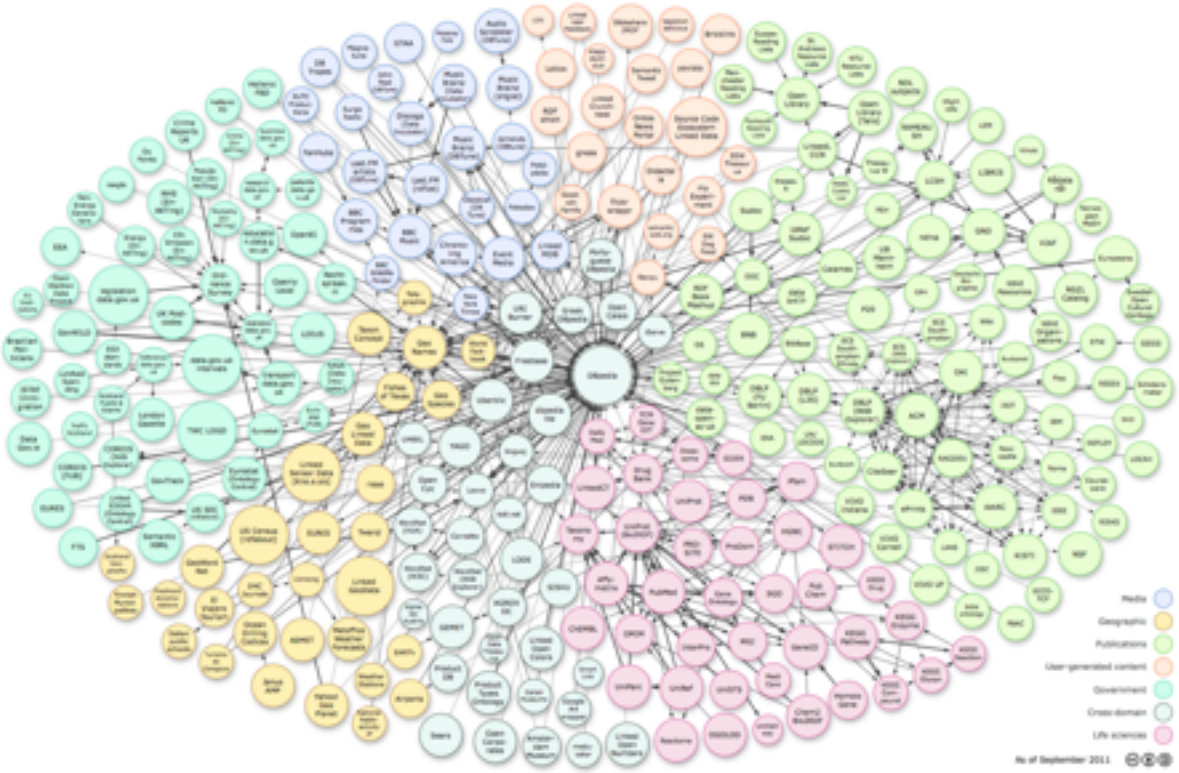

Figure 9: Theme based, Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

The size of each bubble is a rough indication of the amount of data in that data set. Some interesting items are:

- Freshmeat, one of the classic sites that lists open-source data
- MusicBrainz, an online database of digital music tracks and albums
- Project Gutenberg, a venerable initiative to make out-of-copyright texts freely available
- FOAF, an RDF approach to social networking
- DBPedia, an LOD wrapper around Wikipedia articles

Despite its wild success, much can be done to improve the Web. At the heart of the W3C's efforts for doing so is the Semantic Web, which would create a network of semantically transparent data. LOD is basically a very Web-developer-friendly path to the Semantic Web, and one that neatly complements the most important Web 2.0 concepts. For example, take mashups, where you take a service output from site A and mix it into one from site B. With

LOD this doesn't have to be such a conscious process, specialized for each component site. You really get to draw transparently from a wealth of data and services scattered across the Web. Some will be free for use, and some will be restricted for security or commerce, but these are just details that Web developers have already sorted out, for the most part. LOD means making it easier for people to discover important things you place on the Web, and making it easier for them to do unexpected, fruitful things with them. Leaders in modern Web projects should start by thinking in terms of what information and non-information resources are represented in the application, and do everything possible to give each one a well-designed HTTP URI and a semantically rich data format, and create links, links, and more links.

Linked, Managed Data put into practice

The foundations of Linked, Managed Data are identifiers, relationships, policy and services.

Identifiers. The library community has already had a pioneering role in the issue of strong, universal identifiers for bibliographic resources and associated authorities. It remains important to apply this framework to the many entities in the library information management and cataloguing space, rationalizing of identifier systems for books, serials, journals, authors, key service points, credible digital resources, and more. This even applies to less rigorous approaches to identification, such as tag schemes for folksonomies. In some cases such rationalization will be difficult, but some work has already been undertaken such as the establishment of authority files.

Relationships. One of the many factors that led to the success of the Web was in how it defined links in a very simple manner, which encourage cross-referencing. Whereas with identifiers, a system of authority is important, with relationships, success comes with looser coupling. The Web's infamous 404 "Not Found" code at first worried purists of hypertext, but history has proven that the power of Web linking lies precisely in its simplicity at the point of definition. Links gain richness through collective effect, one famous example being Google's PageRank algorithm. The library community can benefit from the same Google-like power of collective Web relationships among key applications.

Policies. Some cultural heritage institutions manage their community with a system of licenses, subscriptions and other factors setting the terms and entitlements for sharing of

information resources, including catalog and authority records. These would have to be incorporated into the metadata framework, through dynamic query, and the conventions set down to assert what processes attach to what resources.

Services. Library applications will bring the ingenuity of the community to bear on cultural heritage systems in innovative new ways, supporting all the many varied interactions between resources allowed on the Web. The use of Web friendly systems for identifiers and relationships, as discussed above, is the most important factor. The metadata framework will also be used for application/service description and discovery.

A challenge we face as a community is in re-thinking libraries and related memory organizations in this context of these foundations.

Related Library Initiatives

Many experts in the Cultural Heritage space have come to the conclusion around the same time that it is important to move towards a Linked Data model. The Library of Congress seeks to increase the chances for interoperability among the various initiatives as they develop. In order to inform efforts in proposing a Linked Data model for MARC, we have researched the efforts and outcomes of several related initiatives that inform the MARC Linked Data model. This section reflects these observations that are based on a combination of personal experiences working with these initiatives, supporting documentation and other communications that these initiatives use to describe their projects.

British Library

Overview / Background

The British Library is developing a version of the British National Bibliography as Linked Open Data. At the time of this analysis the offering includes only published books. Future releases plan to extend coverage to include serial publications, multi-part works, integrating resources (e.g. loose leaf publications), kits and forthcoming publications.

Implementation Status

As of the writing of this report, the work is in progress. The BL has made a dataset available (referenced in the date endpoint section), but makes clear that it does not guarantee the dataset to be stable as its contents may change at any time without warning as a part of ongoing development.

The BL is refreshing the dataset on a regular basis and continuing to work on resolving some issues.

Data Model and Serialization

The British National Bibliography (BNB) of 3.3 million titles was used as the basis for their work. This differs from their first bibliographic publishing initiative using RDF in that they did not set out solely to encode their collection of MARC records into RDF/XML, as they and several other libraries have done previously.

Instead, they modeled 'things of interest', such as people, places and events, related to a book you might hold in your hand. Although they were constantly aware of the value of the BNB data, they did not want to be constrained by the format and practices that went into its creation over many years.

The graphical display of the underlying data model can be found here: http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf

Following Linked Data practices adopted across the web by governments, business and academia, they modeled these 'interesting things', reusing as many existing descriptive schema as possible. There still remained areas for which there was no appropriate existing property with the right meaning. To fill this gap, a set of British Library Terms (BLT) was developed and will be published alongside the data.

The namespace for this schema is http://www.bl.uk/schemas/bibliographic/blterms# This schema, along with the following vocabularies are used to describe the resources of the British National Bibliography (BNB):

- Bibliographic Ontology (Bibo)
- Bio: a Vocabulary for Bibliographical Information

- Dublin Core
- International Standard Bibliographic Description (ISBD)
- Org: An Organization Ontology
- SKOS
- RDF Schema
- OWL
- FOAF
- The Event Ontology (http://purl.org/NET/c4dm/event.owl#")
- WGS84 Geo Positioning and the preliminary version of RDA

The data is connected to various other 'linked open data' sources, in particular: VIAF, LCSH, Lexvo, GeoNames (for country of publication), MARC country and language codes, Dewey.info and RDF Book Mashup.

Data End Points

The subset of the British National Bibliography, covering material published or distributed in the UK since 1950, is currently available at:

- SPARQL endpoint: http://bnb.data.bl.uk/sparql
- Describe endpoint: http://bnb.data.bl.uk/describe
- Search service: http://bnb.data.bl.uk/search and http://bnb.data.bl.uk/items
- The Data Hub: http://thedatahub.org/dataset/bluk-bnb (full triples file)

All files are distributed under a Creative Commons CC0 1.0 Universal Public Domain Dedication license.

Observations

The British National Bibliography (BNB)'s efforts have been very influential in terms of defining the BIBFRAME model. BNB's modeling of subjects in particular has been extremely valuable in terms of generalizing the BIBFRAME Authority model.

Deutsche Nationalbibliothek

Overview / Background

The Deutsche Nationalbibliothek (DNB) offers a Linked Data service that will permit the semantic web community to use the entire stock of its national bibliographic data, including all authority data. The long term goal of the service is to attract new target groups and aim to support the groups' requirements and specific needs. To this end, the DNB have generated documentation to explore the needs of new target groups such as commercial service providers (operators of search engines and knowledge management systems alongside research institutions and non-profit organizations).

The DNB's stated goal is to make an important contribution to the global information infrastructure with its new data service by laying the foundations for modern commercial and non-commercial web services.

Implementation Status

The DNB started to publish its authority data as Linked Data in 2010. The Linked Data Service has expanded and includes bibliographic data since January 2012. As a first step, the bibliographic data of the DNB's main collection and the serials (magazines, newspapers, and series) of the German Union Catalogue of Serials (ZDB) have been converted to a Linked Data format.

Data Model and Serialization

Data available via DNB's Linked Data services are serialized in RDF/XML and RDF/turtle. The bibliographic data records modeled in RDF are currently not of the same complexity as their MARC counterparts. Specific properties based on the type of resource being modeled (periodical, book, collection, article, series, etc.) are utilized. DNB's first phase of implementation has, quite sensibly, been a gradual refinement and expansion of the model.

The choice of vocabularies used for describing DNB resources was made from widely available vocabularies which have been re-used by other stake holders. Because the vocabulary of existing ontologies does not always meet the requirements for representing the full depth of data structures, only existing ontologies were re-used to facilitate rapid and

pragmatic conversion of the bibliographic data. These include a mixing of the following ontologies:

- Bibliographic Ontology (Bibo)
- International Standard Bibliographic Description (ISBD)
- Dublin Core
- RDF Schema
- OWL
- FOAF

Data End-Points

The German National Library provides raw data dumps from their Linked Data services page http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata_node.html . Sample, individual instance data is also available at http://thedatahub.org/dataset/deutsche-nationalbibliografie-dnb. Sample, individual record of the Gemeinsame Normdatei (Integrated Authority File) are available at http://thedatahub.org/dataset/dnb-gemeinsame-normdatei .

The authority files of the German-speaking countries, Gemeinsame Normdatei (GND), and the bibliographic data in the Linked Data service will fall under the Creative Commons Zero (CC0) license with immediate effect.

DDC German will fall under the Lizenz Creative Commons BY-NC-ND license. This applies to resources with URIs beginning with http://d-nb.info/ddc.

Observations

The 'type' based approach to resource description has been incorporated into the BIBFRAME model discussed earlier in this document. DNB's sensible approach to refinement and expansion of the model over time is also reflected in the BIBFRAME design. DNB's authority files for supporting German-speaking countries has been valuable in terms of generalizing the BIBFRAME Authority model.

OCLC/WorldCat

Overview / Background

On June 20, 2012, OCLC updated WorldCat.org to include Schema.org based descriptive mark-up to WorldCat.org pages. This work provides additional linkages among their bibliographic description and other linked data efforts in the authority community (such as VIAF) and subject / taxonomy efforts (including FAST and DDC). With the addition of Schema.org mark-up to all book, journal and other bibliographic resources in WorldCat.org, the entire publicly available version of WorldCat is now accessible to intelligent Web crawlers, like Google and Bing, that can make use of this metadata in search indexes and other applications.

Implementation Status

OCLC has made a commitment to stable, but continued improved functionality of the linked bibliographic data. Their initial effort has focused around leveraging, when possible, schema.org vocabulary and extending this vocabulary only when needed. This initial release is described as experimental and subject to change.

Data Model and Serialization

The Linked Data model for OCLC follows schema.org plus a fledgling library specific extension. Each item in WorldCat has embedded RDFa and microformats data.

Data End-Points

This Linked Data release of WorldCat.org is made available by OCLC under the Open Data Commons Attribution License.

Observations

Zepheira has been involved in the design and development of the OCLC work. The focus of this effort has been to project the bibliographic data available in WorldCat into a schema.org focused vocabulary for supporting increased discovery in schema.org participating search engines. Further, this effort has been a first step in identifying what is needed beyond the schema.org vocabulary for supporting more effective discovery of library specific materials.

The combination of this work, coupled with the historical requirements of MARC (specifically resource description) has contributed to the BIBFRAME model.

Schema.org

While not specifically library related Schema.org is a significant, year-old development for the Web and Linked Data. In 2011 Google, Microsoft Bing and Yahoo, the world's three largest search engines, recently launched an initiative, Schema.org, to allow Web publishers a means to express rich metadata. This supports improved search engine results and is designed to help evolution towards the data Web. Compatibility with Schema.org has become an important way for organizations in many areas of interest to improve the Web presence and network value of their materials.

Libraries will also emerge on Schema.org and there has already been discussion and draft Schema.org modules on how to represent library materials. OCLC has produced an experimental Schema.org-compatible representation of WorldCat, but this work has no recognized standing with the core Schema.org initiative.

Observations

While the business goals for schema.org are not aligned with the objectives of the Library community, the potential impact this work may have on the library community is enormous. While the BIBFRAME model is focused on defining a common model where Library Linked data efforts can interoperate, the design is also influenced by the larger schema.org efforts. A natural outcome of the BIBFRAME work will be to leverage this, and the initial work started by OCLC, to help lead the Library discussions within the schema.org community and shape this work in a way that best supports Library needs.

Resource Description and Access (RDA)

Overview / Background

RDA is the cataloging model that succeeds AACR2. RDA is published jointly by the American Library Association, the Canadian Library Association, and the Chartered Institute of Library and Information Professionals (CILIP) in the UK. Maintenance of RDA is the responsibility of the Joint Steering Committee for Development of RDA (JSC). The JSC is

composed of representatives from the American Library Association, the Australian Committee on Cataloguing, the British Library, the Canadian Committee on Cataloguing, CILIP, Deutsche Nationalbibliothek, and the Library of Congress.

RDA is an application of the FRBR (and FRAD) conceptual models. It is a set of instructions that have been shaped by the FRBR model. In 2009, the three U.S. national libraries (the Library of Congress (LC), the National Library of Medicine (NLM), and the National Agricultural Library (NAL)) agreed to make a joint decision on whether or not to implement RDA based on the results of a test to assure the operational, technical, and economic feasibility of RDA.

Implementation Status

RDA was initially released in June 2010.

In early 2012, the U.S. RDA Test Coordinating Committee recommended that the three U.S. national libraries (Library of Congress, the National Agricultural Library, and the National Library of Medicine) adopt RDA once certain conditions are met based on the testing mentioned in the overview. The implementation will begin no earlier than January 1, 2013. Accordingly, the Library of Congress announced it will have fully implemented RDA cataloging by March 31, 2013. Several other national libraries including the British Library, Library and Archives Canada, National Library of Australia, National LIbrary of New Zealand and Deutsche Nationalbibliothek also plan to implement RDA in 2013.

Data Model and Serialization

The RDA data model is built on FRBR Functional Requirements for Bibliographic Resources, but does not prescribe a particular method of encoding or display of the data. There are mappings from RDA to FRBR, MARC 21 and MODS.

Observations

See FRBR section below.

Functional Requirements for Bibliographic Resources (FRBR)

Overview / Background

FRBR is a conceptual entity-relationship model developed by IFLA. FRBR focuses on user tasks of retrieval and access in online library catalogs and other bibliographic databases. It represents an approach to retrieval and access as derived from the relationships between entities that form links to navigate through the hierarchy of relationships. FRBR focuses on four user tasks to support its model:

- **Find** - to find entities that correspond to the user's stated search criteria
- **Identify** - to identify an entity
- **Select** - to select an entity that is appropriate to he user's needs
- **Obtain** - to acquire or obtain access to the entity described

Implementation Status

There has been a lot of research conducted on how the implementation of FRBR might affect online public access catalogs.

OCLC launched a number of FRBR related research projects with WorldCat records. FRBR has also been implemented in the Australian Literature Gateway (AustLit) with some modifications and extensions to the basic model. The Library of Congress provided a simple tool to convert search result sets to FRBR hierarchical displays.

Several commercial vendors have also designed FRBR implementations. The Virtua catalog developed by VTLS Inc. is marketed as offering full support of the FRBR Model (Virtua 2005), while Portia has created VisualCat, an integrated cataloging system that is capable of consolidating different types of metadata within a single semantic framework based on RDF and FRBR (VisualCat 2005).

Some more recent implementations are noted in the FRBRblog: Library Thing, Drupal, Austrialian Music Centre and the University of Toronto.

Data Model and Serialization

There are three components in an entity relationship model: entities, attributes and relationships between entities.

Group 1: product of intellectual or artistic endeavor (work, expression, manifestation, item)
Group 2: those responsible for the intellectual or artistic content (person, corporate bodies)
Group 3: subjects (concept, object, event, place + all entities in groups 1 and 2)

FRBR is a conceptual model and doesn't define a particular transfer syntax. Several efforts to define XML Schemas for FRBR have been proposed including the Variations/FRBR Project at Indiana University - http://www.dlib.indiana.edu/projects/vfrbr/index.shtml

Observations

The RDA and FRBR efforts have been one of the key contributions in re-focusing cataloging efforts from 'strings to things' and in providing a set of base line functional requirements for supporting the future of cataloging. The holistic approach to retrieval and access as defined by the FRBR work has been a guiding principal to the model proposed in this document.

ONIX

ONIX is a family of standards for communicating detailed metadata about books, serials, and other published media, using common data elements. The ONIX standards include ONIX for Books, ONIX for Serials, and ONIX for Licensing Terms. ONIX for Books is the international standard for representing and communicating book industry product information in electronic form. ONIX for Serials is a group of XML formats for communicating information about serial products and subscription data.

ONIX is developed and is maintained by EDItEUR, jointly with Book Industry Communication (UK) and the Book Industry Study Group (U.S.), and has user groups in Australia, Belgium, Canada, Finland, France, Germany, Italy, the Netherlands, Norway, Russia, Spain, Sweden and the Republic of Korea.

Implementation Status

In April 2009, EDItEUR announced the release of a major new version of the ONIX for Books standard: ONIX 3.0. This release of ONIX is the first since 2001 that is not backwards-compatible with its predecessors and, more importantly, provides a means for

improved handling of digital products. A revised version (3.0.1) was subsequently released in January 2012.

A list of some of the organizations that have implemented ONIX for Books within their business can be found at http://www.editeur.org/111/users-and-services-directory.html. The list includes publishers, retailers, logistics companies, software developers and digital services providers, and represents only a small fraction of the organizations using ONIX.

The implementation of ONIX for Serials is currently unknown to the authors of this report at this time.

Data Model and Serialization

As an XML-based standard, each release of ONIX for Books consists of an XML Document Type Definition (DTD) and/or Schema, together with the associated documentation that specifies the data content of a standard ONIX message or data file. EDItEUR provides these specifications, various XML tools, plus guidance on how to implement ONIX, and use of all these materials is free of charge. O. PNIX is not in itself a database – it is a way of communicating data between databases – but many EDItEUR members and other organizations provide commercial off-the-shelf software or web-based applications for product management that implement ONIX messaging. Other members have developed their own in-house solutions that implement the ONIX communication standard.

The XML DTDs for the ONIX standards are based on the Indecs content model. Indecs was built from a simple generic model of commerce. The top-level model is summarized by Godfry Rust (one of the original architects of Indecs) in the following way "people make stuff; people use stuff; and (for commerce to take place) people make deals about the stuff".

The Indecs model is defined in 4 different conceptual layers:

- **Abstraction** - A creation which is a concept
- **Expression** - An event which is a creation (may be a performance)
- **Manifestation** - An artifact (a creation which is a thing) containing an infixion (or encoding) of an expression (is either physical, e.g., a book, or digital, e.g., an MP3 file)
- **Item** - A single instance of an artifact

There are similarities between FRBR and Indecs however each was informed by different functional requirements, and so evolved different mechanisms for dealing with the issues that seemed most important to them. Each is a particular view on the "universe of discourse" of resources and relationships: there are many valid views. Broadly, they are compatible, and effective integration of metadata from schemes based on them should be achievable, but they must be handled with care.

Observations

The ONIX / Indecs work is related to the BIBFRAME work indirectly though the various alignment efforts between the ONIX efforts and RDA. The ONIX for Serials effort requires further evaluation as it may provide a potential basis for future BIBFRAME modeling efforts for Serials.

Conclusion and Future Directions

The Library of Congress led Bibliographic Framework Transition Initiative is starting the effort to translate MARC 21 to a Linked Data (LD) model. This document outlines an initial model for the interchange of data in a Linked Data environment based on the analysis and synthesis of related activities. To further refine this effort, work continues along the following path:

- Continued analysis of related initiatives
- Vocabulary Navigator for navigating the terms defined for Work, Instance and Authority descriptions and relating these terms to MARC 21 codes
- Elaboration on the XML Serialization of the BIBFRAME RDF model including a sample test suite of instance data
- Tools for supporting the transformation from MARC 21 to the BIBFRAME model
- Linked Data Browsers of BIBFRAME data to help demonstrate the benefit (from a cataloging as well as patron perspective) of representing MARC 21 as Linked Data

Glossary of related terms

HTML (HyperText Markup Language) : the predominant markup language for web pages.

HTTP (the Hypertext Transfer Protocol) : an Application Layer protocol for distributed, collaborative, hypermedia information systems such as the Web.

JSON (short for JavaScript Object Notation) : a lightweight computer data interchange format. It is a text-based, human-readable format for representing simple data structures and associative arrays.

Linked Data : a method of exposing, sharing, and connecting data via the Web architecture.

Linked Open Data (LOD) : a method of exposing, sharing, and connecting freely available data via the Web architecture.

RDFa (or Resource Description Framework – in – attributes) : a W3C Recommendation that adds a set of attribute level extensions to XHTML for embedding rich metadata within Web documents. The RDF data model mapping enables its use for embedding RDF triples within XHTML documents, it also enables the extraction of RDF model triples by compliant user agents.

Representational State Transfer (REST) : a style of software architecture for distributed hypermedia systems such as the World Wide Web.

Resource Description Framework (RDF) : a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax formats.

Resource Description Framework Schema (RDFS) : a family of World Wide Web Consortium (W3C) specifications that builds on the core RDF standards and defines a set of classes with certain properties that provide basic elements for the description of of RDF vocabularies (sometimes referred to as light weight ontologies).

Serialization: The process of converting a data structure or object state into a format that can be stored (for example, in a file or memory buffer, or transmitted across a network connection link) and "resurrected" later in the same or another computer environment.

SPARQL : a query language and protocol for RDF.

Semantic Web : a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF).

Simple Knowledge Organization System (SKOS) : a family of formal languages designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary. SKOS is built upon RDF and RDFS, and its main objective is to enable easy publication of controlled structured vocabularies for the Semantic Web.

Uniform Resource Identifier (URI) : a string of characters used to identify a name or a re-source on the Internet. Such identification enables interaction with representations of the re- source over a network (typically the World Wide Web) using specific protocols. See http://en.wikipedia.org/wiki/Uniform_Resource_Identifier. The purpose of grounding all ontological terms in URIs is to promote re-use across communities of interested parties.

Uniform Resource Locator (URL) : a subset of the Uniform Resource Identifier (URI) that specifies where an identified resource is available and the mechanism for retrieving it on the Web.

World Wide Web Consortium (W3C) : is the main international standards organization for the World Wide Web.

XML (Extensible Markup Language) : a set of rules for encoding documents electronically.

XSLT (XSL Transformations) : a declarative, XML-based language used for the transformation of XML documents into other XML documents. The original document is not changed; rather, a new document is created based on the content of an existing one.