In 2012, Zepheira was engaged by the U.S. Library of Congress to lead the design of a linked data replacement for MARC, the very successful but half-century old library catalog format. We made sure to begin every part of the discussion by considering user stories and the tangible problems they represented, but we already had ideas about the bigger challenges libraries were facing and how our work might help. The result is BIBFRAME, a bibliographic framework meant to provide basic concepts to support flexible metadata for catalogers. BIBFRAME’s ambitious goal is to support the library community of the future, building on lessons learned from the past without being locked into the past.

BIBFRAME builds an architecture on the web, complementing web standards to connect library data with the larger web of data more effectively and efficiently. As we see it, the biggest problem faced by libraries and similar institutions is that they are hardly visible on the web. They are losing public influence and impact rather than providing much-needed leadership to the information age.

At Zepheira, we had already encountered this problem while leading a different project for the Library of Congress. Viewshare.org is a platform in which curators can create special collections of digitally preserved material [1].
Curators at important institutions had been making heroic efforts against the loss of shared cultural heritage, but their results were almost entirely dark to the public. Their organizations already know that they must do a more effective job of putting preserved materials on the web, but too many initiatives in support of this goal get caught up in questions of how best to construct vocabulary for their metadata. A number of comprehensive formats such as RDA (Resource Description and Access) have been developed with great deliberation. There is value in such activity, but there is also a problem when potential users tend to wait for the format to be perfected, slowing down the move to make better use of the web.

The idea behind BIBFRAME is to reduce delay. BIBFRAME provides consistent yet flexible means to communicate data and is designed to make the simple things simple and the complex things possible. Its inspiration lies with the decades-earlier development of the Dublin Core Metadata Element Set (DCMES), which famously favored simplicity in the form of a mere 15 terms. DCMES has been very successful but is a bit too flat to represent more detailed descriptive practices (such as data encoded in MARC). To help address this issue, the Dublin Core Metadata Initiative (DCMI) provided a means of element refinement (DCTERMS). DCMI also introduced the “1:1 principle” (http://wiki.dublincore.org/index.php/Glossary/One-to-One_Principle) that provided a basis for linking together small, descriptive and correlated sets of resource cataloging metadata. BIBFRAME also starts with a small core vocabulary to represent the richer structure needed to encapsulate creative work for cataloging purposes. Communities of BIBFRAME users can build layers on the core to support description of books, periodicals, audiovisual material, digital publications and much more. Even more specialized communities can continue the layering process to address cataloging areas from rare books and manuscripts to medical journals. Finally, the extensibility of BIBFRAME supports entirely novel cataloging needs for works ranging from tweets to software packages.

Another important model that informed BIBFRAME is Functional Requirements for Bibliographic Records (FRBR). FRBR was designed to support access to library materials by distinguishing their more general properties, such as topic, from properties specific to particular editions or physical copies. FRBR has been very successful in getting catalogers to think broadly about emerging modes of user access, such as the web. One difficulty with FRBR is that it creates a universal stratification of creative material that doesn’t always suit the wide diversity of models used to express metadata. This complication adds to general uncertainty and delay as catalogers debate which aspect of FRBR should govern which aspect of a resource to be expressed as linked data on the web. BIBFRAME’s simplified organization around content (Work) and carrier (Instance) derives from lessons learned both from Dublin Core and from FRBR.

BIBFRAME focuses less on the minute details of vocabulary and more on rapid, pragmatic adaptation to what works best on the web. Given the big problem facing memory organizations on the web – their lack of visibility – there is a lot of practical experience from other industries to suggest the solution. Institutions should do much more linking within and across their bodies of useful information. BIBFRAME encourages them to think about the use cases around access in order to optimize how they organize actual
materials, but it also encourages them to perform experiments around linking and description on the web and to rapidly iterate on how they organize metadata.

The things that boost the visibility of sites and pages on the web, such as search engines, indices and viral dynamics in social media, rarely have anything to do with vocabulary. The most important metric in any discussion of visibility is the number and quality of links. Search engine optimization (SEO) is the subject of constant study and even desperate trickery, but the most effective SEO has always been about arranging links between useful content. In the context of visibility there is little point having a page with the perfect vocabulary for some purposes if it is not also part of a rich network of links. BIBFRAME provides a basis for such a network among memory institutions that already have plenty of useful content that is kept in the dark thanks to a poverty of linking.

And that observation brings us to an important clarification. There is some confusion around the Schema.org initiative, launched by the major search engines in order to help content creators provide metadata to improve the functionality associated with search engine results. Schema.org, for example, allows commerce sites to specifically format the price for an item, which can then be displayed right in search engine results for that item, in order to support a use case of comparison shopping. There have been initiatives to refine Schema.org vocabulary sets for bibliographic information. This activity complements rather than competes with BIBFRAME, which can be used to express the relationships inherent in traditional catalog metadata in terms of links. These links enhance the visibility of the page, and once they are visible and appear in search engine results, the value of Schema.org comes into play, enhancing the user’s interaction with those BIFRAME-derived resources through search engine tools.

BIBFRAME is a linked data technology, designed for flexibility, even circumventing some of the purism of other linked data technologies, if these strictures get in the way of the ultimate goals. For example, BIBFRAME can be represented in Resource Description Framework (RDF), and indeed much of the material explaining BIBFRAME uses RDF formats. Yet BIBFRAME is not strictly in line with esoteric details of the RDF model, nor with some of the cherished conventions of RDF formalists. BIBFRAME should be just as straightforward to apply in plain HTML web pages and in dynamic web applications as it is in more highly structured data sets.

BIBFRAME is about four types of things: Work, Instance, Authority and Annotation. Work and Instance have some correspondence to entity types from FRBR and look to separate basic concepts of creative material according to the abstract notion of the creative content (Work class) and the physical objects through which these works are accessed (Instance class). For example, digital bits are considered in the physical domain for purposes of BIBFRAME; therefore an image on the web of Neil Armstrong on the moon is an Instance as is a print photograph that conveys the same picture (which is the Work) to the viewer. Authority provides a way to connect resources to well-managed records of those instances at various institutions, for example, Library of Congress Subject Headings or Getty image identifiers. Annotations provide a mechanism for supplementing the information about a resource, separate from the central characteristics for cataloging that resource. Information about which institution holds a resource or user reviews of a resource can be added through annotations. BIBFRAME recognizes at a fundamental level that the credible resources that are surfaced to the web provide the substrate for further collaboration, organization and enhancement by the users of this data. Annotations provide a means for accelerating this process in a consistent and reusable manner.

FIGURE 1. The BIBFRAME model
Different users and communities can layer additional types of things on BIBFRAME, organized into profiles, which can be conceptually broad (a profile for cataloging print books) or narrow (a profile for cataloging rare books or manuscripts). Profiles can be layered so that, for example, a particular institution’s rules and conventions for cataloging books are preserved to meet use cases around its own unique value.

**Practical Practitioners and a Web of Library Linked Data**

After hundreds of conversations with library professionals, we know this for certain: libraries want to be relevant on the web, and this begins with visibility. BIBFRAME and other linked data formats streamline the integration of libraries into the web so that users can learn what libraries have to offer. At Zepheira we have also found a need to spread knowledge and awareness about how BIBFRAME can practically support relevance and visibility. This need has led us to put our web, linked data and BIBFRAME experience to work in developing a training program focused on libraries, their data, their audiences and their local needs.

It is important for librarians and other metadata professionals to experiment with BIBFRAME using their own data, thinking about their own use cases. We provide utilities for that very purpose, based on our open-source BIBFRAME software, but with some training-specific additions. This process has also created a feedback loop through which valuable and instructive BIBFRAME patterns are captured and fed back to the community, including through the open-source toolkit.

Figure 2 reflects a raw curator’s view of output from Zepheira’s MARC to BIBFRAME transformation service. In this particular case, 11 MARC records are materialized into 212 linked BIBFRAME resources. MARC has a lot of data represented as “strings” which in BIBFRAME come alive as more richly described and organized “things.” Additionally, MARC has defined a large set of contextual relationships (either in the context of indicators and/or subfields), which in a web context, become active links that connect these “things” together. Helping professionals understand this process, and in this context, contribute to the evolution of BIBFRAME has been critical to its utility and performance. The service includes basic search and faceted navigation to support the student in exploration and relating the MARC records to the resulting BIBFRAME resources.

There are also views that illustrate how such data might be presented to the end user, along the lines of traditional discovery, but illustrating how such interfaces can be even richer if not subjected to the limits of legacy records. The first step is to lay out for the cataloger how BIBFRAME resources can be related and...
Rather than trying to perfect their data model, it’s best to proceed through experimentation.

The key to data modeling for BIBFRAME is to recognize that a change in mindset is important in the context of creative works metadata. Most traditional modeling methodologies try to arrange the world into collections of abstract symbols in logical form, but BIBFRAME is not that way at all. In the BIBFRAME mindset the human-readable descriptions of the things being modeled are actually their most important core properties, and the natural associations users might make between one thing and another, however informal or imprecise, are essential for discovery and ultimately organized with different areas of emphasis. Figure 3 is an interface emphasizing BIBFRAME Work resources. Figure 4 emphasizes the corresponding BIBFRAME Instances.

**Linked Data Design from Practice**

After metadata professionals have been exposed to a starting vocabulary based on BIBFRAME and to the possibilities and potential of linking, they can start to develop a picture of how they would adapt BIBFRAME tools and techniques in order to meet their own needs. They can start to assess how ready they are to make their own metadata more visible on the web.
from usage of deployed systems are used for further refinement and iteration, with less frequent change once the service is deployed.

BIBFRAME is designed with all these considerations in mind and also with a mind toward supporting scenarios where SMEs are much more readily available than TEs. SMEs who have gained comfort with BIBFRAME in a hands-on manner, using their own data, can come in on their own key use cases by constructing user stories, which they may scour for nouns and verbs, which they may refine into resources, properties and relationships. The SME can work through such refinement by mapping against BIBFRAME vocabulary registries such as those at the Library of Congress’s bibframe.org (http://bibframe.org/vocab/) or Zepheira’s bibfra.me[c][d][e] (http://bibfra.me/vocab/). In cases where there is an uncertain fit, the SME can determine through testing whether reuse of an existing construct or coinage of a new one works best for the use case, weighing the trade-off of fitness for purpose versus overhead of managing new vocabulary.

The training interfaces described in the previous section illustrate how a BIBFRAME interface can be used to support Interact/Obsenre testing against use cases. There are already many linked data technologies, including web log and dynamic page analytics, that support ongoing refinement of linked data applications, and these technologies are all readily applicable to BIBFRAME.

**Onward to the Visible Library**

All the details of vocabulary and modeling are but theoretical curiosities if the library’s collections are not visible. Visibility through linked data illuminates the richness and value of collections for users who are already familiar with the library, and it also communicates the library’s value to the public, attracting new users. Any effort to represent library information on the web should focus on increasing visibility: vocabulary and modeling details can be refined over time, based on actual analytics of increased usage patterns. BIBFRAME supports such an order of business, building on lessons of the past without too much reverence to the past.

The first generation of BIBFRAME tools, such as those provided through Zepheira’s training, already allows librarians and other metadata
professionals to experiment with models that suit their own institutional needs. These SMEs can then proceed, even if they do not have much technical support, to experiment with publishing linked data in order to increase visibility. As more and more institutions publish linked data and collaborate to increase linking from one collection to another, they will quickly find that they already have everything they need to become the most visible sources of credible information on the web.

Resources Mentioned in the Article