Using BIBFRAME for bibliographic description

Bibliographic description is an essential process of librarianship. In the distant past this process took the form of simple inventories. In the last century we saw bibliographic description evolve from the catalog card to the MARC record. With the advent of globally networked computers and the hypertext transfer protocol, we are seeing the emergence of a new form of description called BIBFRAME which is based on the principles of RDF (Resource Description Framework). This essay describes, illustrates, and demonstrates how BIBFRAME can be used to fulfill the promise and purpose of bibliographic description.

It is asserted here that library collections are not really about books nor any other physical medium because those things are merely the manifestation of the real things of libraries: data, information, and knowledge. Similarly, the primary services of libraries are not really about the lending of materials, but instead the services surround learning and intellectual growth.

Librarianship is now suffering from a great amount of reader dissatisfaction. True, most people believe libraries are "good things", but most people also find libraries difficult to use and not meeting their expectations. People search the Internet (Google) for items of interest, and then use library catalogs to search for known items. There is then a strong desire to actually get the item, if it is found. After all, "Everything in on the 'Net". Right? To this author's mind, the solution is two-fold: 1) digitize everything and put the result on the Web, and 2) employ a newer type of bibliographic description, namely RDF. The former is something for another time. The later is elaborated upon below.

RDF is data model -- a method for organizing discrete facts into a coherent information system, and to this author, this sounds a whole lot like a generalized form of bibliographic description and a purpose of library catalogs. The model is built on the idea of triples whose parts are URIs or literals. Through the liberal reuse of URIs in and between sets of triples, questions surrounding the information can be answered and new information can be inferred. RDF is the what of the Semantic Web. Everything else (ontologies & vocabularies, URIs, RDF "serializations" like RDF/XML, triple stores, SPARQL, etc.) are the how's. None of them will make any sense unless the reader understands that RDF is about establishing relationships between data for the purposes of sharing information and increasing the "sphere of knowledge".

So what does this have to do with libraries and specifically bibliographic description? The answer is not that complicated. The what of librarianship has not really changed over the millennium. Librarianship is still about processes of collection, organization, preservation, dissemination, and sometimes evaluation. On the other hand, with the evolution of technology and cultural expectations, the how's of librarianship have changed dramatically. Considering the current environment, it is time to evolve, yet again. The next evolution is the employment of RDF and linked data as the means of bibliographic description. By doing so the data, information, and knowledge contained in libraries will be more accessible and more useful to the wider community. As time has gone on, the data and metadata of libraries has become less and less librarian-centric. By taking the leap to RDF and linked data, this will only become more true, and this is a good thing for both libraries and the people they serve.

That said, it is in this author's opinion, that the difference between the various ontologies is akin to debating the differences between vanilla and chocolate ice cream. It is a matter of opinion, and the flavors are not what is important, but rather it is the ice cream itself. Few people outside libraries really care which ontology is used. Besides, each ontology includes predicates for the things everybody expects: titles, authors, publishers, dates, notes, subjects/keywords, added entries, and locations. Moreover, in this time of transition, it is not feasible to come up with the perfect solution. Instead, this evolution is an iterative process. Give something a go. Try it for a limited period of time. Evaluate. And repeat. We also live in a world of digital data and information. This data and information is, by its very nature, mutable. There is no reason why one ontology over another needs to be debated ad nauseum. Databases (triple stores) support the function of find/replace with ease. If one ontology does not seem to be meeting the desired needs, then (simply) change to another one. In short, BIBFRAME may not be the "best" ontology, but right now, it is good enough.

Eric Lease Morgan
University of Notre Dame
March 6, 2016

Recipe

Here is a "recipe" for doing bibliographic description using BIBFRAME (or just about any other bibliographic ontology):

1. Answer the questions, "What is bibliographic description, and how does it help facilitate the goals of librarianship?"

2. Understand the concepts of RDF and linked data.

3. Embrace & understand the strengths & weaknesses of BIBFRAME as a model for bibliographic description.

4. Design or identify and then install a system for creating, storing, and editing your bibliographic data. This will be some sort of database application whether it be based on SQL, non-SQL, XML, or a triple store. It might even be your existing integrated library system.

5. Using the database system, create, store, import/edit your bibliographic descriptions. For example, you might simply use your existing integrated library for these purposes, or you might transform your MARC data into BIBFRAME and pour the result into a triple store, like this:

   A. Dump MARC records.

   B. Transform MARC into BIBFRAME.

   C. Pour the result into a triple-store.

   D. Sort the triples according to the frequency of literal values.

   E. Find/replace the most frequently found literals with URIs.

   F. Go to Step #D until tired.

   G. Use the triple-store to create & maintain ongoing bibliographic description.

   H. Go to Step #D.

6. Expose your bibliographic description as linked data by writing a report against the database system. This might be as simple as configuring your triple store, or as complicated as converting MARC/AACR2 from your integrated library system to BIBFRAME.

7. Facilitate the discovery process, ideally through the use of linked data publishing and SPARQL, or directly against the integrated library system.

8. Go to Step #5 on a daily basis.

9. Go to Step #1 on an annual basis.