Attending Theory and Practice of Digital Libraries 2011
Berlin, Germany, September 25-29, 2011
Make Metadata not War

Myung-Ja Han (mhan3@illinois.edu)
Metadata Librarian/Assistant professor of Library Administration
University of Illinois at Urbana-Champaign, USA

1. Introduction

With generous support from Bibliothek & Information International (BI-International) of Bibliothek & Information Deutschland (BID), I was fortunate to attend the 3rd annual ***International Conference on Theory and Practice of Digital Libraries (TPDL)*** (http://www.tpdl2011.org/) held at the University of Humboldt in Berlin, Germany, from September 25th to 29th. As the name of the conference implies, the TPDL has a reputation of being a place where current trends and new research developments of digital libraries are presented and discussed. Because of that reason, people from academia, research institutions, and libraries, who are actively involved in developments and implementations of new technologies for digital libraries, attend this conference to share their experiences and expertise.

Throughout the conference, I learned new trends and technologies in digital library services, and what participating researchers and practitioners are working for their digital libraries. Also, based on what I learned from conference presentations and demos, I came up with ideas that I could implement into my institution's digital library services that would improve user access and experiences. In addition, I met a great group of people with whom I can communicate with on issues and solutions concerning digital libraries and develop a collaborative project. This experience also motivated me to focus on several areas of research that I would like to pursue for myself, such as implementing linked data models into local bibliographic data systems and the provenance of metadata management.

2. Conference

The conference consisted of three parts, doctorate symposium, workshops, and the main conference. Since I only registered for the main conference, from 26th to 28th, the description and observations of this report are focused on the main conference.

The two and a half day conference was composed of three keynote speeches, 12 paper sessions, one panel discussion, and posters and demos. The main theme of

the conference was "*Make Metadata not War*" since metadata is the key to make all digital resources accessible now and in the future. Especially nowadays, libraries are trying to find ways to make their data available to anyone who would like to use it, if the resources don't have copyrights constraints, and to connect to the broader world by providing their data as linked data; thus, the importance of having quality and machine-processable metadata became greater than before.

Because of this reason, the majority of sessions I attended were the ones that discussed how we could enhance metadata, which would ultimately facilitate better resource discovery for data consumers, i.e., regular users as well as data harvesters. The methods that the papers presented could be summarized in three ways:

- using vocabulary mappings,
- implementing FRBR concepts into non-bibliographic data,
- analyzing query terms.

2-1. Vocabulary Mapping

Vocabulary mapping has been discussed as a way to increase the interoperability of metadata and discovery service regardless of the system. Especially when developing a large-scale portal, such as Europeana (http://www.europeana.eu/portal/), a portal for digital contents created by libraries, museums, and archives in European countries, developing a good vocabulary mapping system would enhance the search and discovery of resources. In the paper *Interactive vocabulary alignment* by Jacco Van Ossenbruggen, Michiel Hildebrand, and Victor De Boer, the authors discussed their experiences with vocabulary mapping in Europeana. Ossenbruggen said that the metadata of aggregate collections from many different institutions were often from pre-defined sets of controlled vocabularies, which should be aligned together for better service. He found that in order to align the vocabulary terms from the different sets, developing a smart system was not enough, i.e., human involvement in mapping and vocabulary alignment is inevitable.

2-2. FRBRization

The new content standard Resource Description and Access (RDA), published in June 2010, was based on Functional Requirement of Bibliographic Records (FRBR). Although the underlying concept of FRBR focused on resources held in a library, the data model can be applicable to a broader data world. Two papers, *Linking FRBR Entities to LOD through Semantic Matching* by Naimdjon Takhirov, Fabien Duchateau, and Trond Aalberg and *Supporting FRBRization of Web Product Descriptions* by Naimdjon Takhirov, Fabien Duchateau, and Trond Aalberg, showed that possibility. I hope that, since FRBR has an abstract data model and its semantics are based on bibliographic description, the future research would be more of a collaborative effort between libraries, which would help develop more practical solutions to apply FRBR concepts to other digital contents.

2-3. Query terms – user provided metadata

As a metadata librarian, my main responsibility is not directly related to the interface design or services . However, metadata that I work with has a direct impact on user experience because the system uses them when users enter a search query. In that sense, papers such as, *An Evaluation of Thesaurus-enhanced Visual Interfaces for Multilingual Digital Libraries* by Ali Shiri, Stan Ruecker, Lindsay Doll, Matthew Bouchard, and Carlos Fiorentino; *Query Operators Shown Beneficial for Improving Search Results* by Gilles Hubert, Guillaume Cabanac, Christian Sallaberry, and Damien Palacio; and *One of these things is not like the others: How users search different resources* by Dana McKay and George Buchanan are all practical and educational in their approaches to look at how users search and what terms they use. The paper by Hubert et al. demonstrated a method they used for suggesting new search terms that could improve search results by analyzing other terms commonly entered by users. McKay and Buchanan's paper confirmed that users have different search behaviors depending on the type of resource. These papers suggest that we cannot provide good service without a good understating of users and a single interface cannot satisfy every user. This made me contemplate the need for metadata harmonization.

3. Metadata Interoperability and (Meta)Data Harmonization

Until recently, the library domain and cultural heritage community talked about metadata interoperability when discussing metadata creation and management in the digital library environment. However, as we use different metadata standards to describe different types of resources, different controlled vocabularies, and use different systems and software to store them, the new term "data harmonization" appeared and caught our attention in conjunction with the semantic web and open linked data initiatives. This brings up the question - what is the difference between metadata interoperability and metadata harmonization?

According to Mikael Nilsson, metadata interoperability is "the ability of two or more systems or components to exchange descriptive data about things, and to interpret the descriptive data that has been exchanged in a way that is consistent with the interpretation of the creator of the data." In comparison, Nilsson defines (meta)data harmonization as "the ability of two or more systems or components to exchange 'combined metadata' conforming to two or more metadata specifications, and to interpret the metadata that has been exchanged in a way that is consistent with the intentions of the creators of the metadata." [1]

Starting from the first day's keynote speech by Thomas Hofmann, the topics relating to what data harmonization can do for users and researchers were discussed in many papers during the conference. Hofmann's speech was especially intriguing since he shared how Google has been working on improving user

---

[1] Metadata Harmonization (NISO/DCMI Webinar) - National Information Standards Organization. (2011). Available from http://www.niso.org/news/events/2011/dcmi/metadata

experiences by translating query with a machine. His point was that digital content is created in specific language but can be translated and used for anyone. In order to provide that service, his research team at Google has been working on developing a cross-language retrieval service by translating queries and OCRd text (even photos!). His talk can be summarized as how Google, the biggest portal, tries to unlock the value of digital content and, in order to do so, data harmonization is involved, which is also what all of us who work for digital libraries would like to accomplish.

Conclusion

As a librarian working at a large research academic library, TPDL provided me an invaluable chance to think about the emerging role of academic librarians. Mentioned by Michael Seedle in his opening speech, since libraries have a new role in advancing scholarly communications, librarians should lead the effort in creating new scholarly document models that will accelerate research cycles, expand the impact of research endeavors, and create new forms of interoperable scholarly resources. In addition, libraries and librarians should promote new creative processes and works that would support intellectual freedom in scholarly endeavors across disciplines. All of these works involve working with information life cycles and that requires good, quality metadata management.