The Architectural Touch: Gestural Approaches to Library Search.

LibViz and Metadata

Presented at the 2015 Digital Initiatives Symposium

April 29, 2015

University of San Diego

Susan Luftschein, Archival and Metadata Librarian, USC Libraries
Abstract

LibViz, a gestural search and browse interface for discovery of library materials, will include materials in many formats. These objects have metadata that is one-dimensional. The systems in which this metadata is stored are one-dimensional. LibViz will be three dimensional. How do we apply this one-dimensional metadata to three-dimensional searching? The project will examine the repurposing of existing metadata (metadata management), and how this metadata can or will support the kind of searching we anticipate the project will allow—the ability to find the ofness or aboutness of an object.
LibViz and Metadata

What is metadata?

Metadata is loosely defined as data about data. This definition is so broad as to be generally unusable, which is why when we talk about metadata we are usually referring to descriptive and/or structural metadata. The former is data that describes contents—in libraries that means the contents of books as well as other types of documents and objects; the latter is information about the structure of data, for example, the structure of a MARC (machine readable catalog) record. In the scope of the LibViz project, both will play an important role.

The objects we anticipate including in LibViz will represent many formats—as many formats as we have in Special Collections. Traditional metadata practices include methods of describing (i.e., applying descriptive metadata to) all kinds of objects one would traditionally find in a library. Traditional metadata is one-dimensional—text stored in linear, albeit hierarchical structures. The systems that use this kind of metadata (online public access catalogs, for example) are also one-dimensional. Database structures are one-dimensional (or perhaps two-dimensional—the point is arguable, but they are certainly not three-dimensional). What we hope to do is apply this one (or two)-dimensional metadata to three-dimensional searching.

We also need to consider how we are going to repurpose our existing metadata (metadata management), and how this metadata can or will support the kind of searching we anticipate the project will allow—the ability to find the ofness or aboutness of an object.

Literature review

When I began to think about the role that metadata will play in this project, I realized there were two aspects that I had to consider. The first, and perhaps most important, is the management of our existing metadata. The second is the representation of “ofness and aboutness” or artifactual and evidentiary information.

Accurate and detailed descriptive metadata housed in an appropriate metadata structure is crucial to the process of digitizing, and then finding, special collections materials, and as more materials in library special collections are being digitized, its importance is growing. Much of the literature on the intersection of libraries, cataloging/metadata, and digitization (Deng, 2010; Zentner-Raasch, 2010; Bell, 2012; Laursen, Christiansen, & Olsen, 2012; Ilik, Storlien, & Olivarez, 2014) is focusing on the reuse of descriptive metadata, which makes perfect sense—why reinvent the wheel when we already have perfectly good metadata about our objects in our library catalogs? The question then becomes, how do we get this metadata out of our catalogs with the least amount of intervention and manipulation? Hence metadata management.

A quick review of recent literature reveals a few unsurprising things. First, metadata management is rapidly becoming an important part of the work of catalogers. As new platforms for discovery are created, as new structure standards for metadata are implemented, as more materials are digitized, the need for the management of metadata becomes increasingly important. Second, most of the projects described when discussing metadata management are
attempting to repurpose metadata from one existing system to another existing system. In other words, the platforms and databases in which this metadata was created and currently lives, and into which this metadata is being repurposed, are not new, nor are the structures in use for their creation and reuse. For example, from a MARC-based system, like OCLC’s WorldCat into a Dublin Core-based digital asset management system, like ContentDM. The majority of these kinds of repurposing projects are able to take advantage of the existence of sturdy structural metadata standards, like MARC and Dublin Core.

The second aspect I realized I would need to consider is the role that metadata plays in ofness and aboutness. These concepts have been discussed by library professionals since the 1980s, with the publication of Sarah Shatford Layne’s 1986 article “Analysing the subject of a picture: a theoretical approach”. Recent literature has focused on information seeking behavior in regards to images and their intended use, rather than on subject analysis. See EunKyoung Chung and JungWon Yoon’s “Image needs in the context of image use: An exploratory study”, which also includes a good bibliography of recent studies on image searching, retrieval, and use in digital contexts. There is an inherent two-dimensional bias in studies like Chung and Yoon’s that goes back to Layne, and her source in the work of art historian Erwin Panofsky. The idea of ofness or aboutness is derived from Panofsky’s 1955 *Meaning in the visual arts: papers in and on art history*, in which he describes three levels of meaning: pre-iconographical, iconographical, and iconological. The first two were discussed by Layne as ofness and aboutness in the context of subject cataloging of pictorial works. These concepts attempt to attach to visual images additional levels of analysis and meaning via enhanced description. I have not yet found anyone discussing the importance and description of ofness and aboutness when discussing three dimensional digital surrogates. This raises the question of if and how our descriptive metadata can work in a new context like LibViz.

**What makes metadata for special collections unique?**

Documents (whether two or three dimensional, books, artworks, etc.) in special collections are housed in these departments because, literally, there is something special about them--age, uniqueness, subject matter, etc. Standard descriptive metadata, like the kind found in traditional library catalogs, is usually not sufficient to adequately describe special collections documents. As any rare book cataloger will tell you, in addition to author/title/subject, rare books (as one example) are generally cataloged at a much more detailed descriptive level. They include not only traditional author/title/subject, but also notes on bindings, publication history, previous owners (provenance), pagination, printer’s marks, etc. (see Figure 1). Archival collections include biographical information, provenance, and scope and contents. All of this information provides researchers with additional avenues of discovery for these kinds of rare and/or unique items.
THE ARCHITECTURAL TOUCH: LIBVIZ AND METADATA

Fig. 1 USC catalog record for Hartmann Schedel’s Liber Chronicarum
Arguably, enhanced descriptive metadata does not usually convey the complete “ofness” or “aboutness” of a rare book, as in the example shown in Figure 1. The former refers to factual subject; the latter to meaning. While this terminology has traditionally been used to discuss description of two-dimensional images (photographs or paintings, for example), I would like to use it in the context of LibViz, as our project can provide a way to enhance searching for ofness and aboutness. I am positing that in a case like LibViz, we are attempting to convey the artifactual and evidentiary qualities of the object, qualities that are akin to, respectively, ofness and aboutness, and which are terms more closely associated with archival materials. The meanings are a bit different—artifactual value refers to the item as a whole, and evidentiary value refers to the contents within its larger context. But there is a relationship between these two sets of meaning that three-dimensional representations of items can reveal and solidify.

LibViz will attempt to address the conveyance of these types of meaning. To discover ofness, researchers can read the catalog record; they will get a good sense, in the abstract, of the factual subject of a rare book. If the object has been digitized, they can also look at two-dimensional representations of that object. If the institution has a sophisticated page turning browser and the entire book has been digitized, they can virtually turn the pages, zoom in (if that capability exists), and examine the object closely in a two-dimensional interface. They can, therefore, also discover its aboutness. Or they can take advantage of projects like the Text Creation Partnership at the University of Michigan, itself a project that built upon early OCR initiatives and that attempts to make texts that cannot be OCR’d available to scholars (Martin, 2007). In addition, text mining and the text encoding initiative are making scientific methodologies for textual analysis available to humanities and social science scholars. While this is exciting for many researchers who do textual or visual analysis of rare books and who cannot see the object in person, this kind of interaction still leaves something to be desired when attempting to ascertain the ofness and aboutness of the object in its entirety. By this I mean the object as a three-dimensional entity—looking at more than just individual pages, but at the wholeness of the object. But...do ofness and aboutness, or artifactual and evidentiary qualities, apply to three dimensional, non-pictorial objects? I believe they do. And does metadata, therefore, play a role in this attempt to convey ofness and aboutness? I believe it does.

Arguably, the development of an interactive, three-dimensional search and browse platform for library objects needs to incorporate these concepts of ofness and aboutness. Conceptual mappings based on concepts other than author/title/subject will require a consideration of metadata that incorporates the kind of descriptive metadata on which special collections catalogers rely. In MARC terms, this metadata can be found in the 5xx fields—notes. Rare book catalogers rely heavily on 590 fields—local general notes (Figure 1). These contain information on a variety of topics unique to the item being cataloged. This information is therefore crucial to providing access to the digital surrogate of a library’s unique item. With other kinds of primary source materials, like those found in archival collections, 520 fields—scope and contents notes—also contain information critical for discovery. The contents of these fields are not governed by structural standards, but neither are they governed by descriptive standards (for the most part). While rare book catalogers and archival catalogers use specialized descriptive standards (DCRM and DACS), there are no hard and fast rules for the construction of the data that these fields are designed to house, although there are certainly best practices. So how do we harvest and repurpose a note like this scope and contents note? (Figure 2)
When we began discussions for the LibViz project, our discussions surrounding metadata were limited to “yes, we can use catalog data for this.” As the date for this conference drew closer, I began to think more seriously about how we would use this catalog data. A number of questions quickly came to mind—about the management of metadata, about the structural standards we would use and that the database would rely on, about the interface, about repurposing metadata.

Managing metadata

What kind of metadata will we include? How will this metadata affect the search results? What kind of schema should we use to enable sophisticated search and retrieval, on the order of complex digital humanities projects? How do we manipulate existing metadata to allow users to refine searches? How does traditional metadata support new kinds of searching and/or enhance discoverability? Is our existing metadata therefore relevant? Is MARC metadata appropriate for our needs? Is the MARC standard designed to do what we need it to do? And what exactly is that? Managing this metadata therefore becomes a crucial part of the process of designing this new tool.

Structure
Are we building a new platform or a new type of database and will we then have to develop a new metadata crosswalk?

**Interface**

Are we creating a new digital library collection or are we creating a new interface for digital objects? Or both? What will the new system look like from a metadata perspective?

**Repurposing**

How can existing MARC metadata be repurposed beyond Dublin Core, which we currently use for digital surrogates, to enhance the user experience, especially in a virtual 3d environment? Will the metadata need to be reformatted in its repurposing?

**Conclusion**

I don’t claim to have answers (yet) to any of these questions. What I can be confident of, however, is that we have structures and standards in place for both MARC and non-MARC metadata that will, once the project's platform and interface progress, allow for interoperability that will enhance discoverability for both beginning and advanced researchers.
References


