Project 1: Metadata Analysis of Online Flickr Collection

Erica Krimmel

San José State University

LIBR 282, Prof. Horodyski

28 February 2014

Introduction

Metadata provides details about an information object above and beyond whatever content is inherently present in the object. In doing so, metadata enhances the findability and usability of the object. In this analysis report, I will review my metadata creation process for an online image collection, "Wildflowers of the Central Sierra Nevada," hosted by Flickr. I am an avid amateur botanist, and love taking pictures of flowers, so this project offered me an opportunity to showcase some of my photos and knowledge. To this extent, I decided that my target audience is other amateur botanists who may be unfamiliar with my area and want a guide to local wildflowers. While designing this collection, I focused on effectiveness in several core areas: object (photo) identification, subject vocabulary, and search retrieval. Each of these will be discussed in context below.

To view my collection, please visit:

http://www.flickr.com/photos/117492196@N06/collections/72157641627952115/

Metadata Fields

Designing the metadata fields for my collection was tricky because Flickr records metadata in several different places–image titles, descriptions, tags, maps, embedded (EXIF) fields, dates, and sets. Each of these places is accessed differently, and having never used Flickr before, I was somewhat frustrated when I knew what I wanted to accomplish but had to waste a lot of time figuring out how to do it. Nevertheless, the metadata fields I ended up entering were as follows…

Title. I used this field to record the scientific genus and species (and subspecies or variation, where present) names of each flower. Although many amateur botanists prefer to use more familiar common names, common names are variable. Since "Title" is the most prominent field,

I needed its vocabulary to reflect a stable identifier for my image; scientific names are not only stable, but also interoperable.

Description. Flickr has a text field box called "Description" where users can put whatever information they like. This field is searchable, although not in the same way as Flickr "tags," which can be aggregated for more controlled results. I chose to include my descriptive metadata in this field:

Common Name. Plain English name for the flower provides secondary identification.

Location Description. Each photo is geotagged, but the "location description" field denotes a more specific place that the flower was growing.

Habitat Description. Short explanation of environment where the photographed flower was found.

Associated species. Other plants growing in the same direct habitat as the photographed flower.

Notes. A free field to record extra information about plant characteristics.

Tags. Tags allow users to aggregate photographs within a single user's photostream or across all of Flickr based on keywords. On the larger scale, this metadata represents a folksonomy because there is no vocabulary control; within the context of my own collection, however, I did control the vocabulary to increase the effectiveness of retrieval. I tagged each photo with one or more of each of these categories:

Plant Family. The Latin name of the family that each flower belongs to, e.g. "Asteraceae." Only one tag allowed for this category, because each flower is only part of one family. If other Flickr users have photographs of flowers, plant families is a likely tag, and because the vocabulary happens to be controlled, this is hopefully a very useful

identification category.

Month. The month set that each photo belongs to, either "June," "July," or "August."

Habitat. A one or two word habitat category, e.g. "Meadow," or "High Elevation." Again, with the vocabulary designed to be potentially compatible with other Flickr users' similar tags. Within my collection, this tag allows users to aggregate flowers based on the habitat that they expect to be looking for them in.

Uses. Human uses for the flower. Options include "Edible," and "Medicinal."

Geolocation. Latitude and longitude of the photograph, marked out on a map view.

Date. Day, month, and year that the photograph was taken.

As the Cornell University Library put it in their digital imaging tutorial, knowing your users is as important as knowing your collection (2000-2003). NISO's 2004 framework for metadata recommends focusing on the purpose of the digital collection, as well as the targeted users' needs and behaviors. I tried to keep these principles in mind while designing and entering metadata for "Wildflowers of the Central Sierra Nevada."

Sub-Organization with the Collection

My Flickr collection includes three sets: June, July, and August. These sets represent a sub-level that organizes the collection by month that each flower can generally be found in bloom. My audience can easily narrow their search to the month they will be flower hunting– although flower bloom times can be unpredictable in the mountains, setting the scope as "month" offers leeway while remaining accurate enough to be useful.

Organizing the collection into month sets also allowed me to add metadata to the set itself. I did not take advantage of this feature as much as I could have, but one brainstorm idea is to add commentary for my audience, e.g. recommendations for a few good flower hunting locations in a

specific month.

Automatically Generated & Bulk Metadata

Flickr automatically extracts and generates several types of metadata, some of which are editable and others which are static. **Date uploaded** records the day, month, and year that I uploaded a photo to the Flickr website; this metadata is static. **Camera** is another static field, where Flickr extracts the camera data (in my case, all "Olympus u720SW,S720SW") from each image's EXIF file. **Location** is a derived field from the map location that I specified, and displays only the nearest city, state, and country. **Content type** is an example of an editable automatically generated metadata field. Although the default content type is "photo," I could change it to "screenshot," or "other art," if either was more appropriate.

Automatically generated metadata is crucial to the overall process and survival of a collection's DAM, as human-created metadata can be very expensive. For a platform such as Flickr, automation is difficult because the platform must be relatively neutral in order to accommodate a variety of users, whereas a platform hand-crafted for an individual institution may be able to predict certain metadata field content. Technical metadata, however, is a good opportunity for automatic generation, at least for images, because many cameras already record metadata such as date, camera, megapixels, even location in EXIF files associated with each image file. Technical metadata also happens to be very important; Dale & Waibel (2004), note that "technical metadata assures that the information content of a digital file can be resurrected even if traditional viewing applications associated with the file have vanished."

Similar to automatically generated metadata is bulk metadata, where I can edit the metadata content of many different images at once. I didn't use this feature very much, but did apply the month tags and certain permissions via bulk editing. Had all of the photos been taken

on the same date, using bulk metadata to record the date would be useful. Overall, I think that bulk metadata is most helpful for yes/no or other limited option content. For instance, I changed my geoprivacy and licensing settings for all images in bulk.

Licensing and Legal Information

I appreciate that Flickr uses the Creative Commons licensing schema, which I think is terrific because it's just as easy to make your work freely available as it is to place varying levels of restriction on it. Personally, I am in favor of open source, shared digital objects, so I chose to label all of my photos with an "Attribute" Creative Commons license.

Aside from the Creative Commons licensing, Flickr does not outline specific legal rights metadata. In the Getty's *Introduction to Metadata* text, Whalen recommends recording the following rights metadata: creator, creator's nationality, creator's dates of birth and death, creation date, copyright status, publication status, and rights research date. Flickr does not need all of these fields, as its objects are (supposedly) all controlled by their original owners. Additionally, it is easy for individual Flickr users to incorporate metadata about the creator (themselves) into their user profile.

Conclusion

This project helped me understand metadata from both creator and user perspectives. As a creator, I struggled more with the Flickr platform than my theoretical design. As a test user of my own collection, I found my experience was helpful to add back into the creation process; only through repeated use will all of the intricate metadata bugs work themselves out.

References

Cornell University Library / Research Department. (2000-2003). *Moving theory into practice: Digital imaging tutorial.* http://www.library.cornell.edu/preservation/tutorial/contents.html

NISO Framework Advisory Group. (2004). *A framework of guidance for building good digital collections, 2nd ed.* Bethesda, MD: National Information Standards Organization [NISO].http://www.niso.org/publications/rp/framework3.pdf

Dale, R. and Waibel, G. (2004). Capturing technical metadata for digital still images. *RLG DigiNews*, 8 (5). Retrieved August 11, 2006, from: http://worldcat.org/arcviewer/2/OCC/2009/07/10/H1247256969461/viewer/file1.html

Gill, T., Gilliland, A., and Woodley, M. Baca, M. (ed.). (2008). *Introduction to metadata: Pathways to digital information, Online edition, version 2.1.* Los Angeles: J. Paul Getty Trust / Getty Standards Program. http://www.getty.edu/research/publications/electronic_publications/intrometadata/pdf.html