

Enhancing Cultural Heritage Collections by Supporting and Analyzing Participation in Flickr

Jacob Jett

Center for Informatics
Research in Science
& Scholarship,
Graduate School of Library &
Information Science,
University of Illinois at
Urbana-Champaign
501 E. Daniel St, MC-493,
Champaign, IL 61820-6211
jjett2@illinois.edu

Megan Senseney

Center for Informatics
Research in Science
& Scholarship,
Graduate School of Library &
Information Science,
University of Illinois at
Urbana-Champaign
501 E. Daniel St, MC-493,
Champaign, IL 61820-6211
mfsense2@illinois.edu

Carole L. Palmer

Center for Informatics
Research in Science
& Scholarship,
Graduate School of Library &
Information Science,
University of Illinois at Urbana-
Champaign
501 E. Daniel St, MC-493,
Champaign, IL 61820-6211
clpalmer@illinois.edu

ABSTRACT

Cultural heritage institutions can enhance their collections by sharing content through popular web services. Drawing on current analyses from the Flickr Feasibility Study, we report on the pronounced increase in use of the IMLS DCC Flickr Photostream in the past year, trends in how users are engaging with the content, and data provider perspectives on participation in Flickr through the DCC. In addition to users providing comments and tags for images, they are increasingly integrating historical images from libraries and museums into new digital objects and special collections. Intermediary services can fill a key role in lowering the burden for institutions to engage in Web 2.0 initiatives and broadening public access to cultural heritage content. To extend the scope of the current DCC services, we propose a feedback framework for transferring user-generated information to institutional data providers.

Keywords

Cultural heritage institutions, digital collections, Flickr, IMLS DCC, Web 2.0, digital aggregation, user-generated content, tags.

BACKGROUND

Many cultural heritage institutions are interested in participating in Web 2.0 services to encourage use of their unique digital materials by the broader public and to gather user-generated information about the images in their digital collections. In August 2009 the Flickr Feasibility Study (FFS) began to develop an experimental service for cultural heritage institutions to share their digital content via Flickr, as an extension of the IMLS Digital Collections and Content (DCC) project. FFS worked with five diverse institutions that contribute content to the DCC aggregation,

two large academic libraries, a university archive, a state library, and a small public library. The resulting IMLS DCC Flickr photostream¹ consists of 8 distinct collections with a total of 4,471 digital objects.

Another objective of the FFS was to explore membership in the Flickr Commons, an initiative that began in 2008 as a joint project between Flickr and the Library of Congress to integrate library digital content with existing Web 2.0 communities (Springer et al., 2008). After a series of discussions over more than 20 months, Flickr, a wholly owned subsidiary of Yahoo, Inc., determined that aggregation services like the IMLS DCC do not fit into their model of membership for the Commons.

The DCC team has continued to examine how cultural heritage institutions can take advantage of Flickr to connect with users outside the Commons, where, in fact, there is more freedom to upload photographs for which copyright restrictions may apply. Previously we have reported on the FFS metadata scheme; methods and workflows developed for uploading images; engagement with the library and museum data providers; and user interactions with the collection (Palmer, 2010; Jett et al., 2010).

In the past 18 months we have documented significant growth in the use of the IMLS DCC photostream and ways that users are interacting with the collections. We have also further explored the service roles that can be offered by an aggregator like the DCC to support participation in Flickr. In this paper we report on analyses of the greatly accelerated use of the photostream and interactions with contributing institutions to examine trends, benefits, roadblocks, and services for cultural heritage institutions that wish to participate in Web 2.0 sharing.

Copyright is held by the authors. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted provided that copies are not made or distributed for profit or commercial advantage and that copies include a full citation. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

¹ <http://www.flickr.com/photos/imlsdcc/>

COLLECTIONS & COMMUNITIES ON THE WEB

There are ample reasons for cultural heritage institutions to share their collections via Web 2.0 services such as Flickr (Affleck, 2007; Burgess, 2007; Zarro & Allen, 2010). One obvious advantage is increased visibility with 51 million registered users.² Sharing cultural heritage content through Web 2.0 spaces expands opportunities for institutions and the public to actively use, and reuse, unique library and museum materials. It effectively counteracts the silo effect of limiting access to cultural heritage content to institutional websites or repositories (Zorich, Waibel, & Erway, 2008). Web 2.0 is all about community (Terras, 2011). Institutions contributing to Flickr share more than just their content; they share a web space that helps define both user expectations and the mutual experience of users and institutions interacting with one another.

User interactions on Flickr range from the personal (selecting a favorite image), to the communal (requesting to add an image to a thematic group), and they can add value to content. Zarro and Allen (2010) identified four ways that users contribute to institutional cultural heritage content on Flickr:

1. Expanding on the information already provided
2. Linking images to other resources
3. Providing corrections to the information provided.
4. Curating the image into a new collection

The first three are valuable for institutions aiming to incorporate user-generated data into their systems and have received quite a bit of attention by the digital library community. Connecting images with Flickr groups created by users is another interesting trend that has not yet been widely examined. These integrated collections generally contain “novel, detailed, and niche content with a very specific scope” and are initiated in one of two ways: by proactively participating—joining the group and adding photos to the group photopool, or reactively participating—requesting photos from group administrators to add to group photopools (Terras, 2011).

As documented in our results below, our experiences in the FFS reinforce previous findings outlined above (Library of Congress, 2008; Springer et al., 2008; Zarro & Allen, 2010), but they also add to a more detailed understanding of the levels of exposure and types of user interactions, and how they can benefit cultural heritage institutions participating in Flickr. For example Flickr attracts users with highly specialized interests and expertise who can make important contributions not just to image descriptions, but also to the enhancement of collections.

IMLS DCC PHOTOSTREAM USAGE

As shown in Figure 1, the frequency of views of images in the eight DCC photostream collections has increased

² <http://advertising.yahoo.com/article/flickr.html>

steadily. It dwarfs usage levels of the large DCC aggregated resource, Opening History,³ which consists of more than 1,530 collections. The photostream has averaged 22,892 views per month for the last 14 months, with a high point of 31,604 views from March 16, 2012-April 15, 2012.

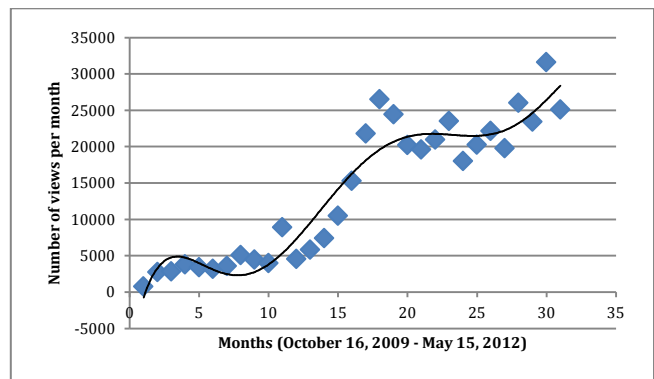


Figure 1. Growth rate in use of IMLS DCC Flickr Photostream over 30 months.

The IMLS DCC participates in 82 groups on Flickr. Thirty-seven of them have directly requested DCC images for reuse. The range of group topics is represented in a phrase cloud in Figure 2. The groups range from very broad collections, such as B&W (Black & White) [photographs],⁴ to very narrow topics, such as Tugs in BC Canada⁵ and Ticket Booth.⁶



Figure 2. Phrase cloud of Flickr Groups that have reused images from the IMLS DCC Photostream.

Participation, especially with the niche, long-tail groups, has been instrumental in defining valuable sub-groups of photographs in existing collections, such as sets of clowns, shipwrecks, and antique vehicles. As emergent specialized strengths are identified in their collections, data providers

³ <http://imlsdcc.grainger.uiuc.edu/history/>

⁴ <http://www.flickr.com/groups/blackwhite/>

⁵ <http://www.flickr.com/groups/1940758@N20/>

⁶ <http://www.flickr.com/groups/ticketbooth/>

can take advantage of the unforeseen opportunities for both collection development and collaboration for building new cross-institutional special collections (Palmer, Zavalina, & Fenlon, 2010).

We are also seeing other novel applications of individual images, such as the example of retrospective photography illustrated in Figure 3. In addition users regularly provide substantive descriptive or contextual information about specific images, through comments, tags or notes (Flickr's method to annotate portions of images), directly enhancing the value of the image within the context of Flickr community spaces. To date Flickr users have enhanced over 300 of the photographs in the DCC Photostream, primarily through commentary. This represents about 10% of the photographs in the photostream, enough to make a valuable contribution to the photostream and a reasonable amount to manage with feedback services, discussed below.

AGGREGATOR AS FACILITATOR

Our interactions with the photostream data providers indicated that technical staff for digital projects can be prohibitively expensive for some smaller institutions. More surprisingly, larger institutions also may not have sufficient personnel to prioritize these kinds of new initiatives (Jett et al., 2010). Thus, as rewarding as engagement with Web 2.0 communities can be, many cultural heritage institutions lack the human resources, expertise, and infrastructure needed to get them off the ground or sustain them once developed. For an individual institution, contributing to Flickr requires personnel time devoted to manually uploading photographs and transcribing metadata, or the technical skill to exploit Flickr APIs⁷ or one of several derived client toolkits.⁸

The Smithsonian achieved local efficiencies in their effort to join the Flickr Commons through coordination among their internal units (Kalfatovic, Kapsalis, Spiess, Van Camp, & Edson, 2009). This case is instructive in many ways, particularly for organizations with large, distributed operations. However, the Smithsonian does not offer an applicable model for implementation of image sharing and other Web 2.0 efforts for most museums and libraries. It is unique in its strong institutional identity, span of administrative structure, and prioritization of digital programs.

The FFS project has shown that coordination through an aggregator is a viable approach for supporting Flickr participation, with moderate cost to individual institutions. An aggregator like the DCC already manages very heterogeneous collections from multiple types of institutions at a national scale and can realize economies of

scale in staff expertise, technical processing, and infrastructure.



Figure 3. Retrospective by Flickr user emanistan⁹ comparing San Francisco Montgomery & Green intersection in 2008 with a Charles Cushman¹⁰ photo from 1953.

For institutions already contributing to the aggregation, resources need only be committed to selection of content and review of metadata. The DCC can provide all uploading services and metadata enhancement, an important feature for collection accessibility (Zarro & Allen, 2010; Kalfatovic et al., 2009), as well as interoperability and functionality. Moreover, as use of the photostream increases and users provide more information, we are developing a service model for providing feedback and analysis of patterns of use to data providers.

A FEEDBACK FRAMEWORK

Since beginning the FFS, the project team has recorded the daily statistics provided by Flickr for views of individual photographs, photostreams, sets, and collections, as well as information about how users were referred to our pages. We also record detailed information regarding user interactions in a Flickr social log that tracks information

⁷ <http://www.flickr.com/services/api/>

⁸ For an example of a client toolkit derived from a Flickr API, see the Sammu tool developed by the Balboa Park Online Collaborative at <http://www.balboapark.org/bpoc/resources/software/sammu>

⁹ <http://www.flickr.com/photos/emanistan/2683587810/>

¹⁰ <http://webapp1.dlib.indiana.edu/cushman/>

such as collection growth, group interactions, user comments, and image re-use. The DCC Feedback Framework recommends a twofold approach to providing feedback to data providers: direct metadata enhancement and data analysis reports.

In the case of the FFS, all participating data providers made their metadata available through the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).¹¹ Using OAI-PMH, we recommend the following workflow for returning enhanced metadata to participating data providers:

1. Record use statistics and user activities.
2. Vet the accuracy of user comments that expand upon or provide corrections to information already provided.
3. Enhance original metadata records with new information.
4. Feed enhanced metadata records to original data providers via OAI-PMH.

The original data providers may then ingest the revised records into their own collections and make the enhanced metadata available for subsequent service providers to harvest. This ensures that item information in other state, regional, and national aggregations will also be updated accordingly. It is important to note that this approach may lead to metadata discrepancies, depending on how often institutions refresh their harvests, and version control is not supported by all metadata standards. In the future, modeling metadata records using an RDF-based standard should provide greater flexibility for curating metadata enhancement and feedback.

Additional user interactions, such as curating images into new collections, re-using images outside Flickr, and linking images to other resources, need to be reported back to institutions manually on a regular basis. The interactions should be quantified and systematically described with an emphasis on identifying niche user groups and hidden collection strengths to inform collection evaluation and development as well as community outreach.

CONCLUSION

As use of the IMLS DCC Photostream has surged, the project has developed a feedback model to guide provision of important usage information back to data providers. The DCC Feedback Framework articulates steps for the collection and quality control of user-generated information, including intelligence on user linking and reuse activities and revealing user-curated emergent collections. Comprehensive feedback of interactions could be further elaborated with longitudinal analysis of patterns and trends in use and comparisons with comparable

institutions to capture the broader dynamics emerging from and around an institution's shared content.

ACKNOWLEDGMENTS

This work was supported in part by Institute of Museum and Library Services, LG-06-07-0020. All expressed opinions are those of the authors, not the funding agency.

REFERENCES

- Affleck, J. (2007). Memory capsules: discursive interpretation of cultural heritage through digital media. (Doctoral dissertation). University of Hong Kong.
- Burgess, J. (2007). Vernacular creativity and new media. (Doctoral dissertation). Creative Industries Faculty, Queensland University of Technology.
- Freeman, C. G. (2010). Photosharing on Flickr: Intangible heritage and emergent publics. *International Journal of Heritage Studies*, 16 (4), 352–368.
- Jett, J., Palmer, C.L., Fenlon, K., & Chao, Z. (2010). Extending the reach of our collective cultural heritage: The IMLS DCC Flickr Feasibility Study. *Proc. Am. Soc. Info. Sci. Tech.*, 47: 1–2.
- Kalfatovic, M., Kapsalis, E., Spiess, K., Van Camp, A., & Edson, M. (2008). Smithsonian Team Flickr: a library, archives, and museums collaboration in web 2.0 space. *Archival Science*, 8(4), 267–277.
- Library of Congress. (2008). On the record: Report of the Library of Congress Working Group on the Future of Bibliographic Control. Retrieved May 21, 2012, from www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf
- Palmer, Carole L. (2010). Aggregating collections to advance our collective cultural heritage. WebWise Conference, March 4, 2010, Denver CO.
- Palmer, Carole L., Zavalina, Oksana, & Fenlon, Katrina. (2010). Beyond size and search: Building contextual mass in aggregations for scholarly use. Proceedings of the American Society for Information Science and Technology Annual Meeting, October 24–27, 2010, Pittsburgh, PA.
- Springer, M., Dulabahn, B., Michel, P., Natanson, B., Reser, D., Woodward, D., & Zinkham, H. (2008). *For the Common Good: The Library of Congress Flickr Pilot Project*. Retrieved May 21, 2012, from http://www.loc.gov/r/print/flickr_report_final.pdf
- Terras, M. (2011). The Digital Wunderkammer: Flickr as a Platform for Amateur Cultural and Heritage Content. *Library Trends*, 59(4), 686–706.
- Zarro, M.A. & Allen, R.B. (2010). User-Contributed Descriptive Metadata for Libraries and Cultural Institutions. *Lecture Notes in Computer Science*, 6273, 46–54.
- Zorich, D., Waibel, G. & Erway, R. (2008). Beyond the silos of the LAMs: Collaboration among libraries, archives and museums. Report produced by OCLC Programs and Research. Published online at: www.oclc.org/programs/reports/2008-05.pdf

¹¹ <http://www.openarchives.org/pmh/>