Learning to Share: Measuring Use of a Digitized Collection on Flickr and in the IR

**Melanie Schlosser
and Brian Stamper**

ABSTRACT

There is very little public data on usage of digitized library collections. New methods for promoting and sharing digitized collections are created all the time, but very little investigation has been done on the effect of those efforts on usage of the collections on library websites. This study attempts to measure the effects of reposting a collection on Flickr on use of the collection in a library-run institutional repository (IR). The results are inconclusive, but the paper provides background on the topic and guidance for future efforts.

INTRODUCTION

Inspired by the need to provide relevant resources and make wise use of limited budgets, many libraries measure the use of their collections. From circulation counts and in-library use studies of print materials, to increasingly sophisticated analyses of usage of licensed digital resources, the techniques have changed even as the need for the data has grown. New technologies have simultaneously presented challenges to measuring use, and allowed those measurements to become more accurate and more relevant. In spite of the relative newness of the digital era, "librarians already know considerably more about digital library use than they did about traditional library use in the print environment."[1] ARL's LibQUAL+,[2] one of the most widely-adopted tools for measuring users' perceptions of service quality, has recently been joined by DigiQUAL and MINES for Libraries. These new StatsQUAL tools[3] extend the familiar LibQUAL focus on users into the digital environment. There are tools and studies for seemingly every type of licensed digital content, all with an eye toward better understanding their users and making better-informed collection management decisions.

Those same tools and studies for measuring use of library-created digital collections are conspicuous in their absence. Almost two decades into library collection digitization programs, there is not a significant body of literature on measuring use of digitized collections. A number of articles have been written about measuring usage of library websites in general; Arendt and Wagner[4] is a recent example. In one of the few studies to specifically measure use of a digitized collection, Herold[5] uses Google Analytics to uncover the geographical location of users of a digitized archival image collection. Otherwise, a literature search on usage studies uncovers very little.

Less formal communication channels are similarly quiet, and public usage data on digitized collections on library sites is virtually nonexistent. Commercial sites for disseminating and sharing

Melanie Schlosser (Schlosser.40@osu.edu) is Digital Publishing Librarian and **Brian Stamper** (Stamper.10@osu.edu) is Administrative Associate, The Ohio State University Libraries, Columbus, Ohio.

digital media frequently display simple use metrics (image views, for example, or file downloads) alongside content; such features do not appear on digitized collections on library sites.

Usage and digitization projects

Digitized library collections are created with an eye toward use from their early planning stages. An influential early CLIR publication on selecting collections for digitization written by a Harvard task force[6] included current and potential use of the analog and digitized collection as a criterion for selection. The factors to be considered include the quantitative ("How much is the collection used?") and the qualitative ("What is the nature of the use?"). More than ten years later, Ooghe and Moreels[7] find that use is still a criterion for selection of collections to digitize, tied closely to the value of the collection.

Facilitating discovery and use of the digitized collection is a major consideration during project development. Payette and Rieger[8] is an early example of a study of the needs of users in digital library design. Usability testing of the interface is frequently a component of site design; see Jeng[9] for a good overview of usability testing in the digital library environment. Increasing usage of the digitized collection is also a major theme in metadata research and development. Standards such as the Open Archives Initiative's Protocol for Metadata Harvesting[10] and Object Reuse and Exchange[11] are meant to allow discovery and reuse of objects in a variety of environments, and the linked data movement promises to make library data even more relevant and reusable in the World Wide Web environment.[12]

Digital collection managers have also found more radical methods of increasing usage of their collections. Inserting references into relevant Wikipedia articles has become a popular way to drive more users to the library's site.[13] Some librarians have taken the idea a step further and have begun reposting their digital content on third-party sites. The Smithsonian pioneered one reposting strategy in 2008 when they partnered with Flickr, the popular photo-sharing site, to launch Flickr Commons.[14] The Commons is a walled garden within Flickr that contains copyright-free images held by cultural heritage institutions such as libraries, archives, and museums. Each partner institution has its own branded space - "photostream" in Flickr parlance - organized into collections and sets. This model aggregates content from different organizations and locates it where users already are, but it still maintains the traditional institution/collection structure. Flickr Commons has been, by all measures, a very successful experiment in sharing collections with users. The Smithsonian,[15] the Library of Congress,[16] the Alcuin Society,[17] and the London School of Economics[18] have all written about their experiences with the Commons. Stephens[19] and Michel and Tzoc[20] give advice on how libraries can work with Flickr, and Garvin[21] and Vaughan[22] take a broad view of the project and the partners.

Another sharing strategy is beginning to emerge, where digital collection curators contribute individual or small groups of images to thematic websites. A recent example is *Pets in Collections*,[23] a whimsical Tumblr photo blog created by the Digital Collections Librarian at Bryn Mawr College.

The site's description states, "Come on - if you work in a library, archive, or museum, you know you've seen at least one of these - a seemingly random image of that important person and his dog or a man and a monkey wearing overalls ... so now you finally have a place to share them with the world!" The site requires submissions to include only the image and a link back to the institution or repository that houses it, although submitters may include more information if they choose. Although more lighthearted than most traditional library image collections, it still performs the desired function of introducing users to digital collections they may never have encountered otherwise.

Clearly, these creative and thoughtful strategies are not dreamed up by digital librarians unconcerned with end use of their collections, so why do stewards of digitized collections so rarely collect, or at least publicly discuss, statistics on the use of their content? The one notable exception to this may shed some light on the matter. Institutional repositories (IRs) have been the one area of non-licensed digital library content where usage statistics are frequently collected and publicized. DSpace,[24] the widely-adopted IR platform developed by MIT and Hewlett-Packard, has increasingly sophisticated tools for tracking and sharing use of the content it hosts. Digital Commons,[25] the hosted IR solution created by Bepress, provides automated monthly download reports for scholars who use it to archive their content. The development of these features has been driven by the need to communicate value to faculty and administrators. Encouraging participation by faculty has been a major focus of IR managers since the initial 'build it and they will come' optimism faded and the challenge of adding another task to already busy faculty schedules became clear.[26] Having a clear need (outreach) and a defined audience (participating scholars) has led to a thriving program of usage tracking in the IR community.

The lack of an obvious constituency and the absence of pointed questions about use in the digitized collections world have, one suspects, led to the current dearth of measurement tools and initiatives. Still, questions about use do arise, particularly when libraries undertake labor-intensive usability studies or venture into the somewhat controversial landscape of sharing library-created digital objects on third party sites.[27] Anecdotally, the thought of sharing library content elsewhere on the web also raises concerns about loss of context and control, as well as a fear of 'dilution' of the library's web presence. "If patrons can use the library's collections on other sites," a fellow librarian once exclaimed, "they won't come to the library's website anymore!" Without usage data, we cannot adequately answer questions about the value of our projects or the way they impact other library services.

Justification for study and research questions

There were three major motivations for this project. First, inspired by the success of the Flickr Commons project, we wanted to explore a method for sharing our collections more widely. An image collection and a third-party image-sharing platform were an obvious choice, since image display is not a strength of our DSpace-based repository. Flickr is currently a major presence in

the image sharing landscape, and the existence of the Commons was an added incentive for choosing Flickr as our platform.

Second, the collection we selected for the project (described more fully below) is not fully described, and we wanted to take advantage of Flickr's annotation tools to allow user-generated metadata. Since further description of the images would have required an unusual depth of expertise, we were not optimistic that we would receive much useful data, and in fact we did not. Still, we lost nothing by asking, and gained familiarity with Flickr's capabilities for metadata capture.

The final motivation for the project, and the focus of the study, was the desire to investigate the effect of third-party platform sharing of a local collection on usage of that collection on library sites. The data gathered were meant partly to inform our local practice, but also to address a concern that may hold librarians back from exploring such means of increasing collection usage - the fear that doing so will divert traffic from library sites. We suspected that sharing collections more widely would actually increase usage of the items on library-owned sites, and the study was developed to explore the issue in a rigorous way. The research question for this study was: Does reposting digitized images from a library site to a third-party image sharing site have an effect on usage of the images on the library site?

About the study

Platforms

For the study, the images were submitted to two different platforms - the Knowledge Bank (KB),[28] a library-managed repository, and Flickr, a commercial image sharing site. The KB is an institutional repository built on DSpace software with a Manakin (XML-based) user interface. Established in 2005, it holds more than 45,000 items, including faculty and student research, gray literature, institutional records, and digitized library collections. Image collections like the one used in this study make up a small percentage of the items in the repository. In the KB's organizational structure, the images in the study were submitted as a collection in the library's community, under a sub-community for the special collection that contributed them. Each image was submitted as an item consisting of one image file and Dublin Core metadata.[29]

The project originally called for submitting the images to Flickr Commons, but the Commons was not accepting new partners during the study period. Instead, we created a standard Flickr Pro account for the Libraries, while following the Commons guidelines in image rights and settings. In contrast to DSpace's community/sub-community/collection structure, Flickr images are organized in sets, sets belong to collections, and all images make up the account owner's photostream. A set was created for the images, with accompanying text giving background information and inviting users to contribute to the description of the images.[30] The images were accompanied by the same metadata as the items in the KB, but the files themselves were higher resolution, to take advantage of Flickr's ability to display a range of sizes for each image. All items in the set were publicly

available for viewing, commenting, and tagging, and each image was accompanied by links back to the KB at the item, collection, and repository level.

The collection

The choice of a collection for the study was limited by a number of factors. First, and most obviously, it needed to be an image collection. Second, it needed to be in the public domain, both to allow our digitization and distribution of the images, and also to satisfy Flickr Commons' "no known copyright restrictions" requirement.[31] This could be accomplished either by choosing a collection whose copyright protections had expired, or by removing restrictions from a collection to which the Libraries owned the rights. Third, the curator of the collection needed to be willing and able to post the images on a commercial site. This required not only an open-minded curator, but also a collection without a restrictive donor agreement or items containing sensitive or private information. Finally, we wanted the collection to be of broad public interest. The collection chosen for the study was a set of 163 photographs from OSU's Charles H. McCaghy Collection of Exotic Dance from Burlesque to Clubs, held by the Jerome Lawrence and Robert E. Lee Theatre Research Institute.[32] The photographs, mainly images of burlesque dancers, were published on cabinet and tobacco cards in the 1890s, putting them solidly in the public domain.

Figure 1. "The Devil's Auction," J. Gurney & son (studio). http://hdl.handle.net/1811/47633 (KB), http://www.flickr.com/photos/60966199@N08/5588351865/ (flickr)

METHODOLOGY

Phases

The study took place in 2011 and was organized in three ten-week phases. For the first phase (January 31 through April 11), the images were submitted to the KB. The purpose of this phase was to provide a baseline level of usage for the images in the repository. In phase two (April 12 through June 20), half of the images were randomly selected and submitted to Flickr (Group A). The purpose of this phase was to determine what effect reposting would have on usage of items in the repository - both on those images that were reposted, and on other images in the same collection that had not been reposted. In phase three (June 21 through August 29), the rest of the images (Group B) were submitted to Flickr. In this phase, we began publicizing the collection. Publicity consisted of sharing links to the collection on social media and sending emails to scholars in relevant fields via email lists. These efforts led to further downstream publicity on popular and scholarly blogs.[33]

Data collection

The unit of measurement for the study was views of individual images. To understand the notion of a "view," we must contrast two different ways that an image may be viewed in the Knowledge Bank. Each image in the collection has an individual web page (the item page) where it is presented along with metadata describing it. In addition, from that page a visitor may download and save the image file itself (in this collection, a JPEG). In the former case, the image is an element in a web page, while in the latter it is an image file independent of its web context. Search engines and other sources commonly link directly to such files, so it is not unusual for a visitor to download a file without ever having seen it in context. In light of this, we produced two data sets, one for visits to item pages, and another for file downloads. Depending on one's interpretation, either could be construed as a "view." Ultimately there was little distinction in usage patterns between the two types of measure.

The data were generated by making use of DSpace's Apache SOLR-based statistics system, which provides a queryable database of usage events. For each item in the study, we made two queries; one for per-day counts of item page views, and another for per-day counts of image file downloads (called "bitstream" downloads in DSpace parlance.) In both cases, views that came from automated sources such as search engine indexing agents were excluded from our counts.

Views of the images in Flickr were noted and used as a benchmark, but were not the focus of the study. Unlike cumulative views, which are tabulated and saved indefinitely, Flickr saves daily view numbers for only thirty days. As a result, daily view numbers for most of the study period were not available for analysis, and the discussion of the trends in the Flickr data is necessarily anecdotal.

RESULTS

At the end of the study period, the data showed very little usage of the collection in the repository. This lack of usage was relatively consistent through the three phases of the study, and in rough terms translates to less than one view of each item per day.

Of the two ways of measuring an image "view" - either by counting views of the web page where the item can be found or by counting how many times the image file was downloaded - there was little distinction. Knowledge Bank item pages received between 5 and 38 views per item, while files were downloaded between 5 and 34 times. Further, there were no significant differences in number of views received between the first group released to Flickr and the second.

	KB Item page views			Image file downloads		
	min	median	max	min	median	max
Group A (images released to Flickr in Phase II)	5	10	35	5	9	25
Group B (images released to Flickr in Phase III)	6	10	38	4	9	34

Table 1. The items in the study are divided into Group A and Group B, depending on when the images were placed on Flickr. This table shows that both groups received similar traffic over the course of the study, with items having between 5 and 38 views in both groups, with a median of 10 for both, and between 4 and 34 downloads, with a median of 9 for both groups.

The items attracted more visitors on Flickr, with the images receiving between 100 and 600 views each. With a few exceptions, the items that appeared towards the beginning of the set (as viewed by a user who starts from the set home page) received more views than items towards its end. This suggests a particular usage pattern - start at the beginning, browse through a certain number of images, and navigate away. A more significant trend in the Flickr data is that most views of the images came after publicity for the collection began (approximately midway through the third phase of the study). Again, the lack of daily usage numbers on Flickr makes it impossible to demonstrate the publicity 'bump,' but it was dramatic.

We witnessed a similar, if smaller, 'bump' in usage of the items in the KB after publicity started. We were also able to identify 65 unique visitors to the KB who came to the site via a link on Flickr, out of 449 unique visitors overall. Of those who came to the KB from Flickr, 31 continued on to other parts of the KB, and the rest left after viewing a single item or image.

DISCUSSION

With so little data, we cannot reliably answer the primary research question. Reposting certainly does not seem to have lowered usage of the items in the KB, but the numbers of views in all phases were so small as to preclude drawing meaningful conclusions.

A larger issue is the fact that much of the usage came immediately following our promotional efforts. This development complicated the research in a number of ways. First, because the promotional emails and social media messages specifically pointed users to the collection in Flickr, it is impossible to know how the use may have differed if the primary link in the promotion had been to the Knowledge Bank. Would the higher use seen on Flickr simply have transferred to the KB? Would the unfamiliarity and non-image-centric interface of the Knowledge Bank have thwarted casual users in their attempt to browse the collection?

The centrality of the promotion efforts also suggests that one of the underlying assumptions of the study may have been wrong. This research project was premised on the idea that an openly available collection on a library website will attract a certain number of visitors (number dependent on the popularity and topicality of the subject of the collection) who find the content spontaneously via searching and browsing. Placing that same content on a third-party site could theoretically divert a percentage of those users, who would then never visit the library's site. The percentage of users diverted would likely depend on how many more users browse the third party site than the library site, as well as the relative position of the two in search rankings. The McCaghy collection should have been a good candidate for this type of use pattern. Flickr is certainly heavily used and browsed, and burlesque, while not currently making headlines, is a subject with fairly broad popular appeal. The fact that users did not spontaneously discover the collection on either platform in significant numbers suggests that this may not be how discovery of library digitized collections works. It is not surprising that email lists and social media should drive larger numbers of users to a collection than happenstance - the power of link curation by trusted friends via informal communication channels is well known. What is surprising is that it was the only significant use pattern in evidence.

The primary takeaway is that promotion is key. If we do not promote our collections to the people who are likely to be interested in them, barring a stroke of luck, it is unlikely that they will be found. Anecdotally, promotional efforts are often an afterthought in digital collections work - a pleasant but unnecessary 'extra.' In our environment, the repository staff often feel that promotion is the work of the collection owner, who may not think of promoting the collection in the digital environment, nor know how to do so. As a result, users who would benefit from the collections simply do not know they exist.

These results also suggest that librarians worried about the consequences of sharing their collections on third party sites may be worrying about the wrong thing. The sheer volume of information on any given topic makes it unlikely that any but the most dedicated researcher will

explore all available sources. Most other users are likely to rely on trusted information sources (traditional media, blogs, social networking sites) to steer them towards the items that are most likely to interest them. Instead of wondering if users will still come to the library's site if the content is available elsewhere, perhaps we should be asking of our digital collections, "Is anyone using them on any site?" And if the answer is no, the owners and caretakers of those collections should explore ways to bring them to the attention of relevant audiences.

CONCLUSION

As a usage study of a collection hosted on a library site and a commercial site, this project was not a success. Flawed assumptions and a lack of usable data resulted in an inability to address the primary research question in a meaningful way. However, it does shed light on the questions that motivated it. Are our digitized collections being used? What effect do current methods of sharing and promotion have on that use?

Librarians working with digitized collections have fallen behind our colleagues in the print and institutional repository arenas in measuring use of collections, but we have the same needs for usage data. In the current climate of heightened accountability in higher education and publicly funded institutions, we need to demonstrate the value of what we do. We need to know when our efforts to promote our collections are working, and determine which projects have been most successful and merit continued development. And as always, we need to share our results, both formally and informally, with our colleagues. Measuring use of digital resources is challenging, and obtaining accurate usage statistics requires not only familiarity with the tools involved, but also some understanding of the ways in which the numbers can be unrepresentative of actual use. The organizations that do collect usage statistics on their digitized collections should share their methods and their results with others to help foster an environment where such data are collected and used.

Next steps in this area could take the shape of further research projects, or simply more visible work collecting usage statistics on digital collections. Of greatest utility to the field would be data demonstrating the relative effectiveness of different methods of increasing use. Do labor-intensive usability studies deliver returns in the form of increased use of the finished site? Which forms of reposting generate the most views? What types of publicity are most effective in bringing users to collections? How does use of a collection change over time? There are also more policy-driven questions to be answered. For example, should further investment in a collection or site be tied to increasing use of low-traffic collections, or capitalizing on success?

Differences in topic, format, and audience make it difficult to generalize in this area, but we can begin building a body of knowledge that helps us learn from each other's successes and failures.

REFERENCES

1   Brinley Franklin, Martha Kyrillidou, and Terry Plum. "From Usage to User: Library Metrics and Expectations for the Evaluation of Digital Libraries." In *Evaluation of Digital Libraries: An Insight into Useful Applications and Methods*, ed. Giannis Tsakonas and Christos Papatheodorou, 17-39. (Oxford: Chandos Publishing, 2009). http://www.libqual.org/publications (accessed February 29, 2012)

2   "LIBQUAL+," accessed February 29, 2012. http://www.libqual.org/home

3   "StatsQUAL," accessed February 29, 2012. http://www.digiqual.org/

4   Julie Arendt and Cassie Wagner. "Beyond Description: Converting Web Site Usage Statistics into Concrete Site Improvement Ideas." *Journal of Web Librarianship* 4, no. 1 (2010): 37-54.

5   Irene M. H. Herold. "Digital Archival Image Collections: Who are the Users?" *Behavioral & Social Sciences Librarian* 29, no. 4 (2010): 267-282.

6   Dan Hazen, Jeffrey Horrell, and Jan Merrill-Oldham. *Selecting Research Collections for Digitization*. (Council on Library and Information Resources, 1998). http://www.clir.org/pubs/reports/hazen/pub74.html (accessed February 29, 2012)

7   Bart Ooghe and Dries Moreels. "Analysing Selection for Digitisation: Current Practices and Common Incentives." *D-Lib Magazine* 15, no. 9 (2009): 28. http://www.dlib.org/dlib/september09/ooghe/09ooghe.html.

8   Sandra D. Payette and Oya Y. Rieger. "Supporting Scholarly Inquiry: Incorporating Users in the Design of the Digital Library." *The Journal of Academic Librarianship* 24, no. 2 (1998): 121-129.

9   Judy Jeng. "What is Usability in the Context of the Digital Library and How Can It Be Measured?" *Information Technology & Libraries* 24, no. 2 (2005): 47-56.

10 "Open Archives Initiative Protocol for Metadata Harvesting," accessed February 29, 2012. http://www.openarchives.org/pmh/

11 "Open Archives Initiative Object Reuse and Exchange," accessed February 29, 2012. http://www.openarchives.org/ore/

12 Eric Miller and Micheline Westfall. "Linked Data and Libraries." *Serials Librarian* 60, no. 1&4 (2011): 17-22.

13 Ann M. Lally and Carolyn E. Dunford. "Using Wikipedia to Extend Digital Collections," *D-Lib Magazine* 13, no. 5&6 (2007). Accessed February 29, 2012. doi:10.1045/may2007-lally

14 "Flickr: The Commons," accessed February 29, 2012. http://www.flickr.com/commons/

15 Martin Kalfatovic, Effie Kapsalis, Katherine Spiess, Anne Camp, and Michael Edson. "Smithsonian Team Flickr: A Library, Archives, and Museums Collaboration in Web 2.0 Space." *Archival Science* 8, no. 4 (2008): 267-277.

[16] Josh Hadro. "LC Report Positive on Flickr Pilot." *Library Journal* 134, no. 1 (2009): 23.

[17] Jeremiah Saunders. "Flickr as a Digital Image Collection Host: A Case Study of the Alcuin Society," *Collection Management* 33, no. 4 (2008): 302-309. doi: 10.1080/01462670802360387

[18] Victoria Carolan and Anna Towlson. "A History in Pictures: LSE Archives on Flickr." *ALISS Quarterly* 6 (2011): 16-18.

[19] Michael Stephens. "Flickr." *Library Technology Reports* 42, 4 (2006): 58-62.

[20] Jason Paul Michel and Elias Tzoc. "Automated Bulk Uploading of Images and Metadata to Flickr." *Journal of Web Librarianship* 4, no. 4 (10, 2010): 435-448.

[21] Peggy Garvin. "Photostreams to the People." *Searcher* 17, no. 8 (2009): 45-49.

[22] Jason Vaughan. "Insights into the Commons on Flickr." *Portal: Libraries & the Academy* 10, no. 2 (2010): 185-214.

[23] "Pets-in-Collections," accessed February 29, 2012. http://petsincollections.tumblr.com/

[24] "DSpace," accessed February 29, 2012. http://www.dspace.org/

[25] "Digital Commons," accessed February 29, 2012. http://digitalcommons.bepress.com/

[26] Dorothea Salo. "Innkeeper at the Roach Motel." *Library Trends* 57, no. 2 (2008): 98-123.

[27] For an example of the type of debate that tends to surround projects like Flickr commons, see http://www.foundhistory.org/2008/12/22/tragedy-at-the-commons/. (accessed February 29, 2012)

[28] "The Knowledge Bank," accessed February 29, 2012. http://kb.osu.edu

[29] "Charles H. McCaghy Collection of Exotic Dance from Burlesque to Clubs," accessed February 29, 2012. http://hdl.handle.net/1811/47556

[30] "Charles H. McCaghy Collection of Exotic Dance from Burlesque to Clubs," accessed February 29, 2012. http://flic.kr/s/aHsjua3BGi

[31] "Flickr: The Commons (usage)," accessed February 29, 2012. http://www.flickr.com/commons/usage/

[32] "The Jerome Lawrence and Robert E. Lee Theatre Research Institute," http://library.osu.edu/find/collections/theatre-research-institute/; "Charles H. McCaghy Collection of Exotic Dance from Burlesque to Clubs," http://library.osu.edu/find/collections/theatre-research-institute/personal-papers-and-special-collections/charles-h-mccaghy-collection-of-exotic-dance-from-burlesque-to-clubs/; "Loose Women in Tights Digital Exhibit," http://library.osu.edu/find/collections/theatre-research-institute/digital-exhibits-projects/loose-women-in-tights-digital-exhibit/. Accessed February 29, 2012.

[33] For an example of the kind of coverage it received, see http://flavorwire.com/195225/fascinating-photos-of-19th-century-vaudeville-and-burlesque-performers (accessed February 29, 2012)