Jason Paul Michel

207 King Library

Miami University

501 East High Street

Oxford, Ohio 45014

micheljp@muohio.edu

Jason Paul Michel is the User Experience Librarian at Miami University Libraries, with primary responsibilities being usability testing, web design and exploration and implementation of emerging technologies. He also serves as the liaison to the Anthropology department. He attained his M.S. in Information and Library Science at Simmons College.

Elias Tzoc

303 King Library

Miami University

501 East High Street

Oxford, Ohio 45014

tzoce@muohio.edu

Elías Tzoc is the Digital Initiatives Librarian at Miami University. He assists with providing access and management to the University Libraries Digital Collections and the Institutional Repository Project. Tzoc is co-leading an LSTA MiniGrant project, researching new access/management technologies for digital collections, and implementing web-customization for DSpace and CONTENTdm.

ABSTRACT

The Digital Initiatives department at Miami University, like most Digital Initiatives and Special Collections, has a large amount of rich digital image collections, stored primarily in a third-party database. Typically, these databases aren't findable to the average web user. It was our desire to expose these collections to the wider web environment. We immediately thought of Flickr, the most prominent website for image discovery and sharing. Our problem was scale; the default uploading interface was too inefficient to push forward with the project. We developed a set of PHP scripts devised to interact with Flickr's API to allow for quick and efficient uploading of images as well as metadata such as titles, subjects, tags, and hyperlinks back to our collections. Once the scripts were in place the uploading process was completely automatic. Miami University Library's Digital Initiatives department has uploaded over 5,000 images using this method and the results have been outstanding. New users are consistently attracted to the content and in about a year there have been nearly 200,000 views of our images. This article will discuss Flickr and the importance of contributing library materials to these types of social media sharing communities; explain the technology and programming involved in the project, and discuss our results and the outcomes of project.

Automated Bulk Uploading of Images and Metadata to Flickr

I. INTRODUCTION

The popularity of social media sites continues to increase, effectively transforming the way users search, share, and interact with digital information. For instance, when searching for images on the web, users will typically end up at one of the well-known image sites such as Flickr (http://www.flickr.com). In fact, even if one decided to start using a specific image domain such as image.google.com, chances are the top results will include images from Flickr.

The same phenomenon has changed the game plan for those in charge of managing image collections. Over the past few years libraries have begun posting multimedia content to social media communities such as Flickr and Youtube (http://www.youtube.com). This content typically falls into three categories: outreach, instruction and collections. The utilization of Flickr to post images of library events is a widespread phenomenon. There are over 300 libraries on Flickr, most of whom are posting outreach materials such as images of the library, library staff and library events.

A much smaller subset of libraries are actually posting digital image collections, such as archival images or special collection images, to Flickr. For example, the Oregon State University Archives has posted over 1,000 historical digitized photographs and postcards (Oregon State University Archives 2010). The Library of Congress is most likely the most prominent library to post content to Flickr. They have uploaded over 8,500 digitized photographs since June of 2007 and have written extensively on the

progress of their project (Hadro 2009). Other libraries to post collections on Flickr are the Library of Virginia, the New York Public Library, the State Library and Archives of Florida, the University of Washington, and more.

In addition to images, libraries are beginning to contribute more and more video content to video sharing sites such as Youtube.com and the lesser known (but aesthetically more clean and professional) Vimeo.com. Most of the video content being published on these sites are outreach videos, but also library instruction videos. Many libraries are uploading video tutorials about the use of library databases, how to search for books, the process of interlibrary loan, etc. The most striking example of this type of video content is being produced by the Arizona State University Libraries. They have a series of short, visually appealing videos entitled *The Library Minute* in which they present basic library instruction in short, one-minute bites (Arizona State University Libraries).

Similar to the digital image trend, a smaller number of libraries and institutions are posting actual video collections on video sharing sites, such as the Library of Congress, the National Archives, and the Vermont Libraries.

In July 2008, the Social Networking Working Group at Miami University Libraries produced a report highlighting emerging technologies that could be used to increase interest in the libraries. One of the technologies investigated was Flickr. The combination of this report and the Library of Congress pilot project inspired a team of two librarians to research and evaluate the feasibility of uploading a set of images from two image collections originally stored in CONTENTdm -Digital Management System, owned by OCLC. The major challenge faced at Miami University Libraries was *scale*.

Using the default upload process in Flickr would have been slow, inefficient, and expensive. Especially because we wanted to include at least three metadata fields (title, subject, and reference URL) for every image, it became obvious that we needed an automated method for a batch upload of images along with three associated metadata fields. The solution was Flickr's API which provides a suite of scripts to allow developers remote access to the data infrastructure.

In February 2009, we uploaded two image collections in Flickr and as of February 2010, we have seen positive results. The image collections have received double the number of views, as well as interaction and feedback from users. Also, the library now has access to item-level statistics. Despite the technical challenges in creating collections with hundreds of items, research studies have shown that Flickr can be an excellent *new* access point to local collections. Saunders (2008) provides a comprehensive overview of the benefits of Flickr for creating digital image collections; he also makes a basic comparison between Flickr and other "formal" digital content management systems. This article will delineate the project from inception to completion and includes a section on the benefits of Flickr for digital collections that may inspire similar projects.

II. BACKGROUND

The Miami University Libraries' Digital Initiatives department aims to preserve and provide easy access to the Libraries' unique and special collections to support student and faculty research both at Miami University and elsewhere. For the Flickr project, we selected two collections:

- **The Victorian Trade Card Collection**, a collection of over 1,500 advertising trade cards from the late 19th and early 20th century from the Charles and Laura Dohm Shields Collection.

- **The Frank R. Snyder Photograph Collection**, a collection of over 4,000 photographs that captures Miami University's early history and life in Oxford, Ohio in the early 20th century and other broad areas of interest including parades, fairs, animals, and sporting events.

Both image collections were originally stored in CONTENTdm version 4.3. During the early analysis of the potential benefits of having a second access point, we made a basic comparison of features available in CONTENTdm and Flickr. The following table is a summary of an initial comparison of ten features in both systems:

[place table 1]

Even though we found some overlap in functionality, this analysis also helped us realize that the new collections in Flickr will automatically add four new features:

- Display of multiple file sizes, useful for presentations or blog posts.

- Support for tagging, which can enable a way to gather "metadata" from regular users.

- Statistics at the item level, which is useful for analysis and administrative reports.

- Support for creating slideshows and sets, a selling point for sharing and interaction.

In February 2009, we launched the Miami Digital Collections project in Flickr and uploaded over 5,500 images and their associated metadata using the upload functionality of the Flickr API (Application Programming Interface) using exported XML

files from CONTENTdm. The same process, however, can be done with any Digital

Management System that allows for XML exporting. Once the upload process and files

are in place, the PHP scripts can fully upload an image and its associated metadata in

approximately five seconds; the batch process, once initiated, can take place with

minimal intervention. Since completing the uploads, our images are viewed around 280

times a day for a total number of nearly 200,000 views.

III. IMPLEMENTATION

After evaluating the series of options available through the Flickr's API, we

decided to use the PHP option, as we felt more familiar with this scripting language.

However, Flickr provides a whole suite of scripts that allows web developers remote

access to Flickr's data and data infrastructure. Another benefit we found in working on

this project is that there is a healthy Flickr community of web programmers,

photographers, and software experts willing to help out with any project. While

investigating the API, we began conversing with a group of people on Flickr's forums

and established a collaboration with a programmer from Spain who had developed a

script that would upload images and metadata from a basic XML file. Since

CONTENTdm has a feature allowing administrators to export customized metadata files

we found this as the best option and we began formalizing a workflow for the upload

process.

The workflow involves four primary elements:

1. **The API key and secret code**, needed to authenticate connectivity between
    local computer and Flickr servers.

2. **Image folder**, a local public web directory where all images should be stored.

3. **The XML file**, a file based on the Flickr XML template that includes all metadata –image filenames must match the filenames in the image folder.

4. **PHP scripts**, files required to establish and validate connectivity and parse XML file into Flickr; note, the upload process could be done in any web programming language.

If you would like to implement this system for your collection, follow along with these steps. All of the files and scripts you will need can be downloaded here: http://beta.lib.muohio.edu/~jpmichel/flickr_test/flickr_uploader.tar.

[place figure 1]

3.1 API Key

Before you begin the process you will need to obtain an API key and secret code. Go to http://www.flickr.com/services/api/misc.api_keys.html to begin the process. You will need to give the application a name of your choosing. Next, select "Web Application" as the app type. Finally, enter the URL to auth.php, the authentication script included in the suite of scripts. Be sure to save your API key and secret code as you will need it later in the process.

3.2 The Images

To make things simple, create a public web directory entitled flickr_uploader or something similar. In this directory, create a subdirectory entitled images. This is the

directory where you will place your images for the upload process. Nothing should be in this directory except the image files, preferably in .jpg format.

3.3 The XML File

In the description section of each image in the Flickr interface, users have the ability to add text and basic HTML code that usually help describe an image; we decided to include the Subject (TGM) terms and create a hyperlink of the Reference URL that links back directly to the corresponding image in our local CONTENTdm collection. Additionally, since the subject (TGM) field contained a list of terms from a controlled vocabulary and they were often used to find/browse similar items, we decided to use them as separate tags in Flickr. In the tag element, we also added some new metadata at the collection level, e.g. Miami Digital Collections.

Example of a Flickr XML format

```
<?xml version="1.0" encoding="utf-8"?>
<collection>
  <record>
    <title> Miami women's basketball team 1907</title>
    <description>
    Subject (TGM): Sports; Universities and colleges; Women's education;
    Basketball players; Basketball uniforms; Coaches (Athletics);
    Group portraits;
    Persistent URL: http://digital.lib.muohio.edu/u?/snyder,4215
```

</description>

<tags>

Sports; Universities and colleges; Women's education;

Basketball players; Basketball uniforms; Coaches (Athletics);

Group portraits; Miami Digital Collections

</tags>

<file>2063842134.jpg</file>

</record>

</collection>

[place figure 2]

3.3.1 Transforming XML files

Most *digital management systems* provide ways to export metadata, in this case, we used the tab-delimited option in CONTENTdm. Ideally, we could have exported an XML file, however, in version 4.3 the filename was only exported using the tab-delimited option. Once the metadata (export.txt) file is in place, follow these steps:

▪ Open this file (in MS Excel) and select fields with meaningful data, e.g. Title, Subject (TGM).

▪ Perform a quality control check of filenames against the real images.

▪ Save new file with only selected fields.

- Run a PHP script that transforms the tab-delimited file into the new Flickr-XML format.

  This is where we created an XML <record> element for every image. The title remained the same; the major change was in the <description> element, here we decided to include the data from Subject (TGM) and URL – if we needed to include all the metadata fields, we would have added them here; the next <tags> element contains the data from subject (TGM) but Flickr will handle each keyword as a separate tag, this is where we added some new tags at the collection level; finally, the <file> element which should match the image file names.

[place figure 3]

After running a few tests using the PHP script, it became clear to us that in order to let others use our script, we needed to create an online form where users can upload a tab-delimited file (based on our example) and create a new XML file; this form also gave us a the ability to add new metadata (tag) at the collection level. A copy of the script is available at: http://adler.lib.muohio.edu/~tzocea/files/flickr/

[place figure 4]

3.4 PHP Scripts

There have been many API kits designed using different programming languages. The class that we used for this process was phpFlickr written by Dan Coulter (Coulter 2010). Working with this class our upload script does the following:

1. Determines the authenticity of the API key and secret.

2. Creates a connection between your data and Flickr's database infrastructure.

3. Works iteratively through and uploads your images and metadata stored in the XML file to the Flickr profile to which you are logged in.

In order to complete this process for your collection, you will only need to edit two php files. First you will enter your API key and secret into line 11 of upload.php and secondly you will do the same into lines 9 and 10 of auth.php. Once your images and XML file are in place you just need to point your browser to upload.php and the upload process will begin.

IV. BENEFITS

4.1 Benefits of Flickr – Findability, Sociability, and Shareability

The primary benefit of posting your digital images to flickr is the wider audience your collection will enjoy. Your collections will be *findable.* Hosting your images on your own insitutions' web site is an important thing to do but it does not ensure that your content is getting as much exposure as it potentially could. The web is an unimaginably large place with many back alleys (including most library Web sites) and a few popular thoroughfares (such as Flickr). If possible it is best to put your collections in a place where many web users reside.   In addition to exposing your collection to a wider

audience, Flickr has many other affordances, including a personal and social interaction with users, increasing shareability of content and insight into users search behavior.

Unlike many institutional library web sites, Flickr has a built-in comment feature, allowing any logged-in flickr user to post a comment about any image. This feature allows you to connect personally with users and makes the images more approachable and accessible. Figure 5 shows an example of the types of comments we receive about our collections. By posting your images to Flickr, you are also increasing the shareability of the images. In most institutional web sites, the images are locked down, but in Flickr, any image can be automatically emailed, posted to a blog, tagged, embedded, and shared. This shareability is extremely important because the more that your images are shared and posted to other places on the web, the more likely that it will be found via search engines. The shareability of the images will increase their findability.

[place figure 5]

4.2 Administrative Affordances

On the administrative side Flickr is very user-friendly and intuitive. Once you uploaded your images you can edit them in batches: add tags, descriptions, and more. Additionally you can drag and drop your images to create specialized sets. Other useful administrative features in Flickr (with a Pro account) include: rotating images, replacing images while keeping the title, description, comments, notes, tags, groups and favorites associated with their original photo. For collection administrators, perhaps one of the most important features in Flickr is the statistics view option; this is where you can:

- Have a quick overview of the activities of today, yesterday, last week and all time

- discover a list of your most popular images,

- classify your images marked as favorites or with comments,

- learn about how people are finding your images,

- find the exact number of views in a given time period (usually daily, weekly or monthly),

- have a list of referrals, e.g. images.google.com.

- use Flickr apps to generate interactive reports, e.g. Flickrstats (Vispillo.org 2010).

Keeping track of the number of visitors to a site —or collecting data about how these visitors found your site—can be useful for several reasons: evaluating, reporting, and planning. This option has allowed us to incorporate a monthly report along with other digital collection's statistical reports. This approach has helped us compare results and confirm that since January 2009 we've doubled the number of views (from ~250 to ~500 views per day). Also, Flickr has become consistently one of the top ten referral sites bringing web traffic to our local CONTENTdm collections.

V. LESSONS LEARNED

After the first year of using Flickr, we're excited to share some positive lessons we have learned; based on results and usage experience we know that Flickr can:

- *be an excellent secondary access point for local image collections*: when users search for images on the web, they most likely go to sites like Flickr or use the image subdomain of major search engines; images in Flickr will often rank at the top in any of those cases.

- *enable interaction intuitively*: once users are logged in they can tag, comment, favorite, annotate, or create new/external sets of images; this type of functionality is still very limited in major Digital Management Systems. http://www.flickr.com/photos/ohiohistory/galleries/72157623147409228

- *help expose collections to non-target audiences*: as one of the most popular sites on the web, images in Flickr will automatically increase visibility to broader groups of users; for instance, back in October 2009, we had a formal request from an institution from the UK.

- *provide affordable online collection alternative*: for only $24.95 a year, anyone can create a Pro account and have unlimited uploads, storage, sets, collections, statistics on your account, and ad-free browsing pages.

For administrative tasks, two possible problematic areas (also learned) in Flickr are:

- *all-time statistics***:** although Flickr used to provide access to the "all time" statistics, that changed in March 2009. As of March 2010, they're letting users download the statistics as *.csv for November 2008 to present, but this

option will only work until June 1 2010. Options like this are very useful when generating annual or semester reports.

- *controlled vocabulary*: sometimes we need to update an existing tag (metadata term) for a set/collection of images; without a controlled vocabulary, this is not something that can be done with a couple of clicks or a global update.

VI. CONCLUSIONS

Flickr's popularity has continued to grow and in many respects is the primary web site for image sharing and discovery. As of March 2010, it is the 21st most-trafficked site in the United States. In the current information environment, where information channels are often shared and interconnected among multiple sources, it is imperative to have as many different access points as possible to library collections and materials. After the first year, we have learned that the two collections in Flickr has not only increased 10% of web traffic to our local CONTENTdm site, but it has also enabled new sharing and interaction opportunities with a much broader and global user community. Figure 6 illustrates two examples of questions and suggestions that we have received, in part because we chose to create a second access point where "today's visitors" are instead of requiring them to come and find our local collections.

[place figure 6]

A basic summary of the first year statistics:

Total images	5618	100%
Photos in groups	71	1%
With views	5,433	97%
With comments	114	2%
Favorited	692	12%

View counts (all-time)

Photos and Videos	142,558
Photo stream	28,312
Sets	13,895
Total	184,765

For an institution considering a similar project, the benefits will probably be the same or even better—depending on the scope of the collection. Perhaps the only challenge is the normalization of the metadata files from a local system into the Flickr XML template. As for the upload files, we are happy to share the PHP scripts we used and as we mentioned earlier, the Flickr API library has several options and they support different programming languages.

For us, we will continue to explore new ways of re-using the rich and detailed usage data we gather from Flickr. Right now, we are periodically revising the keywords people have used to find our images and we have been re-indexing those terms in the HTML tag element in local files as well as in the autosuggest search feature on the splash

pages of the Victorian Tradecards and Frank Snyder collections. An automated script that can collect this type of data will be very useful. Another experiment we will begin working on very soon is the ability to batch update the metadata in the description filed. For instance, if we migrated our local collections into a different system, we would like to keep the URL reference in the records in Flickr. Lastly, we will be happy to evaluate the use of the "machine tags" (Flickr.com 2010) in these collections.

VII. REFERENCES

Arizona State University Libraries. 2010. The Library Channel.

http://www.vimeo.com/librarychannel (accessed June 3, 2010).

Coulter, Dan. 2010. phpFlickr. http://phpflickr.com/ (accessed June 3, 2010).

Edmunson-Morton, T. 2009. Talking and Tagging: Using CONTENTdm and Flickr in

the Oregon State University Archives. *The Interactive Archivist.*

http://lib.byu.edu/sites/interactivearchivist/case-studies/flickr-at-osu/ (accessed

June 3, 2010).

Flickr.com. 2010. Flickr API / Discuss.

http://www.flickr.com/groups/api/discuss/72157594497877875/ (accessed June 3,

2010).

Hadro, Josh. 2009. LC Report Positive on Flickr Pilot. *Library Journal* 134(1): 23.

Hadro, Josh, ed. 2009. The Library of Congress Talks Digital Initiatives. *Library Journal*

134(14): 13.

Kalfatovic, Martin, Effie Kapsalis, Katherine Spiess, Anne Camp, and Michael Edson.

2009. Smithsonian Team Flickr: a library, archives, and museums collaboration

in web 2.0 space. *Archival Science* 8(4): 267-277.

Oregon State University Archives. 2010. Oregon State University Archives photostream.

http://www.flickr.com/photos/osucommons/ (accessed June 3, 2010).

Saunders J. 2008. Flickr as a Digital Image Collection Host: A Case Study of the Alcuin

Society. *Collection Management* 33(4): 302 – 309.

Springer, Michelle, Beth Dulabahn, Phil Michel, Barbara Natanson, David Reser, David

Woodward, and Helena Zinkham. 2008. For the Common Good: The Library of

Congress Flickr Pilot Project. http://www.loc.gov/rr/print/flickr_report_final.pdf (accessed June 4, 2010).

Vispillo.org. 2010. Flickrstats. http://flickr.vispillo.org/ (accessed June 7, 2010).