


Performance Report Cover

1. Federal agency and organization element to which report is submitted Institute of Museum and Library Services		2. Federal grant or other identifying number assigned by federal agency #LG-46-13-0257-13		Page 1	Of 1 Pages
				3a. DUNS number 04-399-0498-0003	
				3b. EIN 53-0196584	
4. Recipient organization (name and complete address, including zip code) The George Washington University 2121 I Street, NW, Ste. 601 Washington, DC 20052				5. Recipient identifying or account number 35221	
6a. Project/Grant period start date (mo/day/yr) 09/01/13	6b. Project/Grant period end date (mo/day/yr) 08/31/14	7. Reporting period end date (mo/day/yr) 08/31/14		8a. Final Report? <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No	
8b. Project URLs if any https://github.com/gwu-libraries/social-feed-manager/				9. Report frequency <input checked="" type="checkbox"/> annual <input type="checkbox"/> semi-annual <input type="checkbox"/> quarterly <input type="checkbox"/> other If other, describe	
10. Please transmit the performance report as instructed by the Institute of Museum and Library Services.					
11. Other attachments <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No Transmit accompanying documents with this cover as instructed by the Institute of Museum and Library Services					
12. Certification: By submitting this report I certify to the best of my knowledge and belief that this information is correct and complete for performance of activities for the purposes set forth in the award documents.					
12a. Typed or printed name and title of authorized certifying official Sylvia Ezekilova Assistant Director, Sponsored Projects Administration				12c. Telephone (area code, number, extension) 202-994-6255	
				12d. Email address osr@gwu.edu	
12b. Signature of authorized certifying official 				12e. Date report submitted (mo/day/yr) 12/19/14	
13a. Typed or printed name and title of Principal Investigator/Project Director Daniel Chudnov Director, Scholarly Technology, GW Libraries				13b. Telephone (area code, number, extension) 202-994-0684	
				13c. Email address dchud@gwu.edu	
14. Agency use only					

Burden Estimate and Request for Public Comments: Public reporting burden for this collection of information is estimated to average ten minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the Institute of Museum and Library Services, Chief Information Officer, 1800 M Street, NW, 9th Floor, Washington, DC 20036-5802, and to the Office of Management and Budget, Paperwork Reduction Project 3137-0029, Washington, DC 20503.

Final Report for IMLS Grant #LG-46-13-0257-13

Collecting Twitter: A New Stewardship Capacity

George Washington University Libraries

Award Amount: \$24,550; Matching Amount: \$27,586

Total Expenditure, Award: \$24,209.77; Total Expenditure, Matching: \$27,156.85

Grant timeline: September 1, 2013 - August 30, 2014

Project Director: Daniel Chudnov, Director of Scholarly Technology, GW Libraries
<dchud@gwu.edu>

Project Summary

In our project, we sought to enhance Social Feed Manager (SFM), a prototype application developed at the GW Libraries in 2012 for collecting social media data from Twitter to meet a variety of academic research, teaching, and library collection development needs. The prototype automates the collection of data from Twitter's application programming interface (API) and makes it possible for scholars, students, and librarians to identify, select, collect, and preserve Twitter data for research purposes at little to no marginal cost. It quickly proved to be useful in supporting a diverse set of researchers and students on campus and drew interest from colleagues at peer institutions hoping to achieve similar goals. We solicited input from these colleagues about how we might improve the application to meet a broader range of use cases at multiple institutions. Based on this feedback, we developed and released a series of enhancements to the application, including extensive documentation to introduce new users to SFM and to support systems staff in SFM installation and day-to-day use.

This application and our project address several challenges facing research libraries and many of our peer cultural heritage institutions. Library professionals have long invested time, training, and services into minimizing the drudgery of access to information. Whether this comes through centralizing acquisitions and subscriptions, having ready access to collection stacks and reference services, or enabling online access to discipline-specific databases, they all serve similar goals in assisting members of the communities we serve to get to the information they need. In the case of researchers and students looking at social media sources, we learned that there is a divide between individuals empowered with the

necessary skills and tools to harvest and process data from APIs for themselves (largely in Computer Science and related fields) and a wide range of specialists and students for whom acquiring, transferring, transforming, and preserving a large volume of data from an API is a great challenge. First, then, the SFM application helps us connect these researchers and students with the data they need to conduct their research, enabling them to move quickly from selection and collection to applying research methods using the tools they know best. As a result, we have defined a new way to extend the traditional role that library staff play in supporting access to information, and the successful implementation of this service has helped us to develop new relationships on campus as word of our application spreads among faculty and their students. Finally, we now understand that the use case of access to these same materials for future researchers -- in the mode of special collections and archival materials -- shares several technical aspects with our current application service but requires additional planning and policy work.

In this way, then, we have confirmed that the service we are providing with SFM is of great value to members of the GW community and helps to enhance the role of the library in the research lifecycle. Furthermore, we have verified with colleagues at peer institutions that they see similar needs in preserving and offering access to social media collections to the communities they serve and they also seek to enhance the role they play in their environments through providing similar services. According to feedback from our peers, our work on SFM has helped many institutions moved toward implementing similar services.

Process

We were able to perform the work of this project largely according to the work plan we laid out in our original proposal. The work took place in three primary modes. The first mode was software development, which continued throughout the project timeline. We established a weekly development meeting, defining new milestones, discussing individually ticketed enhancements and bugfixes, and balancing ticket assignments among team members. Our development team comprised (as planned) a project director/manager, a software developer, an e-resources content manager, and a graduate student developer. The full-time staff contributed portions of their time and the graduate student worked twenty hours per week on the project from early in the timeline through the end. During the period of the grant, we released six new versions of SFM (all available under a free/open source software license), roughly one every two months, each marked by a milestone on the project's Github page (<https://github.com/gwu-libraries/social-feed-manager/milestones?state=closed>).

The second major activity of the grant was hosting a two-day meeting at the GW Libraries in December 2013 focused on the application and the broader space of potential services and use cases envisioned by the collective organizations represented at the meeting. We invited interested colleagues from several peer institutions and were able, with the help of project funding, to support the participation of attendees from New York University, the University of North Texas, North Carolina State University, the University of California at San Diego, the University of Arizona, George Mason University, the Digital Public Library of America, Yale University, the University of Virginia, a Presidential Innovation Fellow, and officers of both IMLS and the National Science Foundation. This meeting enabled us to gain a deeper understanding of the needs that our peers had already discovered for applications like SFM, as well as the services we might build around them. It also offered us an opportunity to define the major categories of activity and enhancements that we should focus on during the course of our project. In short, we came to understand that we can classify most potential uses of SFM as serving either the immediate needs of researchers and students or the future needs of researchers and students, which are challenging to anticipate. In the first case, we could note patterns in our use of SFM in helping people to move forward with their work and enhance the application to better meet their needs, particularly with some forms of self-service functions (eliminating the need for staff intermediation). In the latter case, we recognized a need to consider policy and workflows for establishing collection strategies and processing collected data through archival workflows all the way through integration of description and access. Overall, it became clear that SFM itself needed much more extensive documentation to support anyone wishing to implement it locally, as well as supporting high-level introductory materials to introduce the application both to new users and organizational decision-makers in support of strategic program planning at their institutions.

The last category of grant activity involved reporting on our project to the broader

community and soliciting feedback from peers who had attended the December meeting at GW, one of our public talks about the project, or had otherwise connected with us about their potential use of SFM. We were able to share our progress through a presentation at the Coalition for Networked Information (CNI) Fall Membership Meeting in December 2013 (slides available at <http://www.slideshare.net/dchud/capturing-the-ephemeral-collecting-social-media-with-social-feed-manager>), an unconference session on social media archiving at IMLS WebWise in February 2014, a hands-on pre-conference workshop at Code4Lib in March 2014, a panel entitled "Social Media, Archiving, and Preserving Collaborative Projects" at the National Digital Stewardship Emerging Trends Symposium in April 2014, the IMLS Focus Meeting in April 2014, and a paper on the technical challenges of the project in *Code4lib Journal* published in October 2014 (available online at <http://journal.code4lib.org/articles/10097>). The CNI talk was particularly well-attended and well-received; we heard many favorable comments from attendees and many suggestions for future work. At the Code4lib pre-conference session, we helped a number of attendees to install SFM successfully and discovered many opportunities to improve our documentation along the way. For feedback on the application, we reached out to colleagues from 10 organizations already using or planning to use SFM and surveyed their progress using the application, successes and barriers they encountered along the way, and their ideas for how to make the application better and more useful. A brief summary of their responses is included in the next section.

As noted above, we were able to accomplish the stated objectives in our proposal without major changes to our original work plan. Expenses for the project were primarily allocated to two major areas: travel support for guests and other ancillary costs associated with our December 2013 meeting and wages for our graduate student programmer. Grant funds were also used to purchase computing infrastructure and to support conference travel for project staff (Code4lib 2014, the IMLS Focus meeting in NYC in May 2014, and the Society of American Archivists 2014 meeting), enabling us to present our work to diverse audiences. Our total direct and cost-shared project expenses nearly matched our project budget.

Project Results

We consider this project to have been successful in meeting our established goals. We were able to improve the SFM application as befit both growing demand for the service we provide with it at GW Libraries and input we received from colleagues at many organizations. In particular, we are proud to be able to report that colleagues from a wide range of institutions were interested enough in the application and its potential growth and utility for their respective organizations that they agreed to attend our December 2013 meeting. All involved agreed that empowering researchers to collect social media data is a valuable endeavor that aligned well with our existing missions, values, and collection development strategies. In particular, libraries and archives were deemed to be well-suited to perform this service with tools like SFM, whether in support of present-day or future scholarship. The opportunity to establish and enhance relationships with researchers in our communities through the use of services like SFM is important because it can reinforce or re-establish the role of librarians and libraries in the academic research lifecycle.

At our meeting in December 2013, we were struck by three core themes that emerged. First, among attendees there was a definite sense that our organizations need to be performing this work, based on existing experience and various levels of success and failure in this realm. A programmatic distinction also became clear: the use cases that resulted from an effort to respond to present-day research needs and those meant to prepare collections for future scholarship have very different implications for program and service development in our institutions. This became evident in the contrasts between stories told by colleagues from the University of Virginia and New York University, who had experienced some level of success in preserving social media and web-based materials related to recent events, and stories told by GW faculty we had previously supported, who explained how the ready availability and features of SFM had made it possible for them to perform a research study previously deemed impossible, as the application saved them considerable time and effort in their immediate work plans. Through this, we learned together that most of the technical work required to support both the immediate- and future-service models is the same, and as such, SFM itself can and does already support both. However, we will need to do more work to examine archival policies and workflows from acquisition and legal transfer through integrating description and access to materials with existing systems and applications. In addition, some yet-to-be-developed technical features, including the ability to capture "both sides" of a conversation (both one individual's tweets and those of other Twitter users tweeting back to that individual, for example), will be more important in a special collections context.

A second theme that arose from our meeting was the wide array of legal concerns that readily appear when stepping beyond the simple boundaries of collecting modest sets of data for scholarly use by individual researchers. Most social media platforms have strict guidelines forbidding republishing, and each has its own legal statements regarding who "owns" the data each service collects from and provides back to individual account holders. We agreed to

set aside legal discussions in an effort to maintain our primary emphasis on the software and its future enhancement, but we also agreed that there is a clear and pressing need for more institutions to dig into the legal ramifications of expanding the scope of collection development and access to include social media data sources.

The third theme that emerged from our meeting was a need for a focused effort to expand documentation for the SFM application. From providing "elevator pitch" summary materials to share for institutional planning and marketing purposes to detailing the precise aspects of daily system operations, everyone in attendance agreed that GW staff would do well to focus a substantial portion of its grant project efforts on improving documentation. To this end, we devoted one entire release to developing extensive documentation; this is now available both as part of the version controlled, free/open source software repository on Github and in an easily readable format at <http://social-feed-manager.readthedocs.org/>.

In addition to the results we obtained from the meeting, we conducted a series of ten interviews in August 2014 with representatives from the University of North Texas, the University of California at San Diego, Duke University, New York University, North Carolina State University, Stanford University, the University of Virginia, Pennsylvania State University, the University of Indonesia, and the University of California at Riverside (comprising attendees of both our December 2013 meeting at GW and our March 2014 workshop at Code4lib 2014, as well as other institutions with whom we had established discussions throughout the project) to gauge their level of interest and success in using SFM and delivering similar services on their respective campuses. A few summary points from these interviews follow:

- Eight of the ten groups we spoke with had at least installed the application successfully; half of these have continued to use it to collect a modest volume of materials; all who have installed it consider it an experimental or research project rather than a production service
- Several interviewees commented on issues relating to installing, managing, and upgrading the application; although the improved documentation helps a great deal, many institutions have server operating system requirements that differ from GW's or would prefer to see a simpler installation and maintenance process
- Many interviewees indicated that working with the application has highlighted a need to better understand how implementing a social media data collection program integrates with their existing offerings and collections; determining who needs to be involved, where "buy-in" needs to come from, and how they can manage maintenance and support of the application with busy, limited staff are key questions to answer
- Several interviewees expressed an interest in more "user stories" - at GW, we tend to tell stories of collecting data from members of Congress and the news media because those are the uses our early faculty users found valuable; they are looking for use cases focused on authors' papers, university archives, and thematic capture relating to local and national events, among others

- There seems to be some agreement that SFM, as a server-based application that can handle multiple functions simultaneously, has some advantages over simpler command line tools and also does some things more thoroughly than existing hosted operations like Archive-It, so it is therefore a valuable addition to what's already available
- Several interviewees expressed an interest in working out the relationship between SFM and other tools like Archive-It and general web harvesting tools, other platform-specific tools like North Carolina State University's Lentil, and approaches to integrating collected media from multiple platforms (e.g. Twitter and Instagram and Tumblr and web sites) to improve coherence among collections; this is an important area for many of us.

Although the project term has ended, we continue to develop the SFM application and provide services to our community with it, as more and more faculty and students connect with us every month for help in collecting and accessing data. We have received a steady enough flow of requests that we have implemented a tracking system to ensure that we follow through on every request. We are also considering ways to add more self-service functions to the application itself so that more users will be able to collect and transform data for themselves. In addition, we see a continuing need for better guides and documentation for using server-based applications like SFM as well as easier-to-use tools that still require a certain degree of technical facility not readily available to researchers and students who have never learned programming.

Strategically, we recognized a need to develop policies and examine workflows for the creation and augmentation of archival collections comprising social media data. As a result, we applied for and were awarded a grant from the National Historic Publications and Records Commission to begin a new project aimed at this purpose. With the new grant, we will expand SFM to support data from Tumblr and Flickr and we will work with archivist colleagues at GW and partners at other institutions to test out using SFM to integrate the collection of social media with our respective archival workflows.

We believe these are still early days, but that the need we perceived two years ago has been confirmed and a great deal more work is necessary to put our organizations into a position where we can collect from social media sources before platforms disappear or disallow collection. There is considerable room for more organizations and more funding agencies to take on more projects in this space. We are pleased to be on the leading edge of this curve and are very grateful for the opportunity that IMLS provided for us to pursue this project. We are excited to hear from so many colleagues who are also recognizing the needs and issues in this domain as strategic opportunities. We would welcome a greater strategic focus from agencies like IMLS on developing tools, programs, workflows, and policies for capturing and preserving web-based materials; a corollary need that develops soon after collection becomes possible will be to develop new generations of descriptive strategies and access tools for large-scale (hundreds of millions of individual items like tweets) collections of digital media.

Resources

(Multiple authors.) Social Feed Manager software repository, online at Github:

<https://github.com/gwu-libraries/social-feed-manager/>.

Chudnov, Daniel, Daniel Kerchner, Ankushi Sharma, and Laura Wrubel. "Technical Challenges in Developing Software to Collect Social Media Data." In Code4lib Journal 26, October 21, 2014, online at journal.code4lib.org/articles/10097.

Chudnov, Daniel, Bergis Jules, Daniel Kerchner, and Laura Wrubel. "Capturing the Ephemeral: Collecting Social Media and Supporting Twitter Research with Social Feed Manager." Presented at CNI Fall 2013 Membership Meeting, December 2013, Washington DC, slides online at www.slideshare.net/dchud/capturing-the-ephemeral-collecting-social-media-with-social-feed-manager.

Kerchner, Daniel, Daniel Chudnov, Ankushi Sharma, and Laura Wrubel. "Social Feed Manager Documentation." Online at http://social-feed-manager.readthedocs.org/en/m5_002/.

Wrubel, Laura. "GW Libraries: Social Feed Manager." Online at <https://library.gwu.edu/scholarly-technology-group/social-feed-manager>.