Final Performance Report
Narrative: Getting Found

Authors: Kenning Arlitsch, Patrick OBrien, Jean Godby, Jeff Mixter, Jason A. Clark, Scott W.H. Young, Devon Smith, Doralyn Rossmann, Leila Sterman, Angela Tate, & Mary Anne Hansen

This is the final report narrative from November 30, 2014.

Made available through Montana State University’s ScholarWorks scholarworks.montana.edu
Final Performance Report Narrative

“Getting Found: Search Engine Optimization for Digital Repositories”

IMLS Award Number: LG-07-11-0345-11
November 1, 2011 – October 31, 2014

PI: Kenning Arlitsch, Dean of the Library, Montana State University
Co-PI: Patrick OBrien, Semantic Web Research Director, Montana State University

Key Personnel:
Jean Godby, Senior Research Scientist, OCLC Research
Jason A. Clark, Head of Library Informatics & Computing, Montana State University
Scott W.H. Young, Digital Initiatives Librarian, Montana State University
Jeff Mixter, Research Support Specialist, OCLC Research

Other contributors:
Devon Smith, Consulting Software Engineer, OCLC Research
Doralyn Rossmann, Head of Collection Development, Montana State University
Leila Sterman, Scholarly Communication Librarian, Montana State University
Angela Tate, Program Coordinator, Montana State University
Mary Anne Hansen, Research Commons Librarian, Montana State University

Project Goals and Objectives
Goals and objectives from the proposal aligned along three general tracks:

1. Expand search engine optimization (SEO) research.
2. Make recommendations for SEO and publish these as a toolkit. The toolkit will include metadata transformation mechanisms and tools for monitoring and reporting.
3. Disseminate findings and provide training to the community.

Summary of Accomplishments

- Achieved measurable improvements to visibility of digital repositories
- Created SEO Toolkit to measure and monitor digital repository performance
- Developed a data model for institutional repositories using Schema.org
- Developed a citation matching process for structured metadata for IR
- Tested Social Media Optimization techniques
- Conducted research resulting in new directions and future deliverables
- Communicated achievements in numerous publications and presentations
THE FOLLOWING PAGES DESCRIBE ACCOMPLISHMENTS FOR ALL THREE YEARS OF THE GRANT.

Introduction
The research we proposed to IMLS in 2011 was prompted by a realization that the digital library at the University of Utah was suffering from low visitation and use. We knew that we had a problem with low visibility on the Web because search engines such as Google were not harvesting and indexing our digitized objects, but we had only a limited understanding of the reasons. We had also done enough quantitative surveys of other digital libraries to know that many libraries were suffering from this problem.

IMLS funding helped us understand the reasons why library digital repositories weren’t being harvested and indexed. Thanks to IMLS funding of considerable research and application of better practices we were able to dramatically improve the indexing ratios of Utah’s digital objects in Google, and consequently the numbers of visitors to the digital collections increased. In presentations and publications we shared the practices that led to our accomplishments at Utah.

The first year of the grant focused on what the research team has come to call “traditional search engine optimization,” and most of this work was carried out at the University of Utah. The final two years of the grant were conducted at Montana State University after the PI was appointed as dean of the library there. These latter two years moved more toward “Semantic Web optimization,” which includes areas of research in semantic identity, data modeling, analytics and social media optimization.

Deliverables

Deliverable #1: SEO Toolkit (Getting Found Web Analytics Toolkit)
We developed a toolkit to help libraries establish baseline measurements of the SEO performance of their digital repositories. The toolkit includes everything necessary for implementing a Google Analytics dashboard that continuously monitors SEO performance metrics relevant to digital repositories. The toolkit will soon be made available in a Council on Library and Information Resources (CLIR) Report and on its Vimeo channel.

Baseline Measurements
The toolkit offers a process to inventory and coordinate hardware systems, online services such as institutional Facebook accounts, and web properties such as various digital repositories or collections. It helps identify stakeholders as well as the various roles people play and the skills required to implement and maintain a library SEO effort.
Dashboard
Based on Google Analytics, the dashboard allows library staff and administrators to monitor common baseline performance metrics relevant to library repositories. It further enables administrators to ask pertinent questions when changes are observed. For instance, a sudden drop in the number of items indexed by Google in a given collection could prompt an administrator to question the appropriate manager or staff person responsible for that collection.

Toolkit Abstract
Implemented and managed correctly, Web Analytics is an invaluable tool for helping library administrators and staff make informed decisions about internal resource allocations and demonstrate the value of library services to external stakeholders.

Our goal for this "Cookbook"\textsuperscript{1} is to provide a series of basic recipes to integrate the necessary analytics for consistently evaluating and monitoring both general and specific library objectives as a single library institution. Each recipe includes a list of "ingredients," such as emails, forms, spreadsheets, dashboards, and other information necessary to implement a baseline SEO metrics relevant to library administrators and collection managers. Each recipe also includes short “How To” videos that cover the following areas.

1. How to use the Getting Found Web Analytics Cookbook
2. How to Start an SEO Program
3. How to create a Master Google Account for the Library Institution.
   a. How to Improve Google Account Security
4. How to Activate, Configure and Maintain
   a. Google Analytics v3 Web Service
   b. Google Webmaster Tools Web Service

These first recipes are critical and form the foundation for establishing the organization as a single semantic entity in the Linked Open Data cloud with baseline analytics relevant to a diverse library customer base. Google Analytics can be set up and maintained by one or more individuals, depending on the organization’s goals and resource constraints. To improve flexibility and accommodate more libraries, we have defined library SEO Roles that most library organizations considering Web Analytics will have already, formally or informally, incorporated into one or more individuals’ job responsibilities.

We recommend the “kitchen” have a single individual accountable as the Analytics Program Lead. This person should know the customers who want web analytics information, and he/she should have the necessary backing from library Administrator Sponsorship to set up and run a Web Analytics kitchen. This program

\textsuperscript{1} Google’s web services are a complex set of systems designed to collect and report information that improves marketing insight and business profitability. The majority of Google features and capabilities require time, resources and a level of technical marketing sophistication that most libraries do not have. We have used a “cookbook” metaphor to aid communication with typical library staff.
lead will interpret and modify the recipes provided in the Getting Found (GF) "Web Analytics Cookbook" to develop processes to configure the kitchen and build a basic Analytics menu that works locally. The GF "Cookbook" includes recipes for producing solutions that we believe should be found on every library Analytics menu.

**Deliverable #2: IR data model**

In 2014 we developed a model for institutional repositories (IR) that focuses on modeling the various types of materials typically found in IR as well as the various types of entities that are related to the materials. We used popular vocabularies (Schema.org) to model high-level entities and relationships to develop the Resource Description Framework (RDF) ontology. Using existing Schema.org vocabularies helps ensure interoperability of the model both within the Library Science domain, and more importantly, on the Web.

The model was developed with instance data from the Montana State University ScholarWorks IR as reference. Using the sample data, we developed a rich vocabulary that allows for materials, people, universities, colleges, and departments to be identified and connected in semantically meaningful ways. This model is currently being tested against the ScholarWorks dataset. The RDF data that is generated by the model will then be used to conduct sample SPARQL queries that can demonstrate how RDF data can be used to extract unique and important analytics. These analytics can help university administrators highlight academic strengths as well as identify areas where future research and collaboration can be most beneficial.

The need to make IR more visible is a direct reflection and reaction to the changing ways in which people search for and interact with library resources. The effort to use Schema.org as the primary vocabulary helped us align the Montana ScholarWorks IR metadata with popular and web-focused vocabulary standards and also provided us with a data model that is understood and consumed by the major search engines. We were able to successfully map the theses and dissertations metadata into Schema.org and, when needed, supplement existing Dublin Core fields with terms we created as part of an extension vocabulary for Schema.org. The extension terms followed the same standards and practices as those in Schema.org and every attempt was made to position extension terms as sub-classes or sub-properties of existing Schema.org terms. After we finished developing the extension vocabulary we successfully demonstrated how to:

1. Integrate the vocabulary into the existing metadata
2. Use OpenRefine to clean-up and add semantic richness to the data.

The modeling and its implementation using existing metadata served as a proof of concept that this type of work is possible and is consumed by commercial search engines. Google has crawled and indexed the pages, and it is recognizes the structured data that is encoded in RDFa.

To test the RDFa markup, the sample pages were run against both the Google Rich Snippets Tool and the W3C RDFa Validator. It demonstrated that Schema.org can be
used to describe institutional repository content with the precision necessary for a subject matter expert to find what they seek, and at the same time provide a layer of abstraction that allows commercial search engines to help a novice discover new information. This evidence supports the efforts already conducted by OCLC to use Schema.org markup to describe materials accessible from WorldCat.org. The use of Schema.org not only provides a gateway to search engine consumption and indexing but also provides a way for libraries to easily and effortlessly share their data on the Web with groups and organizations outside of the library domain. Conversely, adopting a Schema.org vocabulary will also make it much easier for libraries to use and leverage data published outside of their domain of practice. Using Schema.org vocabulary will help library organizations increase the amount of structured data that they share with search engines. Additionally, the use of Schema.org will allow organizations to connect isolated data silos to the rapidly expanding scope of the Semantic Web to improve discovery, access and value of their intellectual output.

**Deliverable #3: Citation Parsing and Matching Process**

The problem faced by many IR is that their rich metadata is locked in single string citations. This causes problems when there is a need to provide data aggregators such as Google Scholar with metadata that is parsed into specific fields, such as author, article title, date, volume, issue number, etc. In order to make this metadata more accessible and consumable by aggregators, a method is needed for converting string citations into parsed metadata fields.

In our 2012 journal article titled “Invisible Institutional Repositories: Addressing the Low Indexing Ratios of IRs in Google Scholar” we demonstrated that IR are not being harvested and indexed by search engines very well, and in particularly not by academic search engines like Google Scholar. We showed in our paper that one of the main reasons for this problem is that the metadata schema used by many IR is Dublin Core (DC), and because DC does not have a field for each part of a citation Google Scholar guidelines state that Dublin Core should be used “only as a last resort.” Scholarly search engines, (i.e., Google Scholar, Microsoft Academic Search, etc.) must be able to parse a citation and deliver it in any style users want, and since citations in IR metadata tend to be entered as text blocks in the dc:source or dc:citation field search engines can’t read or comprehend the textual strings.

We demonstrated through a pilot study that using metadata schema recommended by Google Scholar (Highwire Press, PRISM, ePrints, or BePress) results in greatly improved harvesting and indexing. The problem now is that IR administrators can’t easily convert their Dublin Core metadata to one of the above-mentioned schema, and there is currently no automated method for doing that conversion, which in turn means that there will likely be no large-scale improvement. To help remedy this problem we have been working on automated citation parsing processes using open source software. To date we have developed a prototype methodology that incorporates citation matching and metadata extraction from OCLC MARC records. This methodology could bring the citation conversion problem for most IR down to a manageable number that have to be converted manually.
Methodologies:

Citation Parsing
The initial idea for converting string citations into individual metadata fields was to use a citation parsing service. There are a variety of open source services available that use entity recognition algorithms to parse citations and convert them into fielded metadata. For the purposes of this study we chose to evaluate three different citation parsing services. Each of the services was tested and then the best was chosen for a production style test of converting 750 citations compiled from Montana State University faculty resumes. The citation service was evaluated based on how many citations were parsed and then a second phase of evaluation was conducted to test the completeness and accuracy of the parsing service.

Citation Matching
After the citation parsing service was tested, a second method for citation parsing was developed and evaluated. This method involved matching the citation to a MARC21 record and then extracting metadata from the record to produce fielded metadata. The algorithm matched strings within the citation to strings in the MARC21 record. A combination of OCLC databases was used to compile a comprehensive set of records to match against. The set included traditional bibliographic records, article records and records from institutional repositories, harvested through the OCLC Digital Collections Gateway. Once the string citations were matched to a specific OCLC record, the MARC21 record was then used to create a fielded metadata record for the item. The services were tested using a set of 100 citations from the University of Utah. In order to evaluate the citation matching methodology, two metrics were tested. The first was how accurately the matching algorithm worked. The second round of evaluation looked at the quality of the extracted metadata after a match was determined.

Results
The initial results of the citation parsing services were encouraging but upon further review, we found numerous errors that required manual correction. Of the 750 citations that were run against the FreeCite service, 457 were positively identified as having been parsed. The additional 293 were identified as not parsed and consequently the fielded metadata was only a best guess. Although this initial statistic seemed to indicate that ~61% of the citations were accurately parsed, we decided to conduct a more detailed analysis of the fielded output in order to make an accurate judgment of the service. Two random samples of 100 citations were selected for meticulous manual review. With the first set, we asked an IR cataloger from Montana State University to manually parse the citations into fielded metadata. We then used the parsed citations to generate metadata for the second set of 100 and asked the same cataloger the go through and clean up/correct the parsed data. The manual parsing took 13 hours while the time it took to clean up the algorithmically parsed data was 20 hours. This suggested that it is more time intensive to clean up metadata generated by the parsing service then it is to simply create the metadata by hand.

The citation matching and extraction process was successful and demonstrated how fielded metadata could be assembled from citations without having to rely on the use
of a citation parsing algorithm. We are expecting that improvements to the algorithms will help increase both matching accuracy and metadata extraction. The graph below shows the frequency of extracting metadata for the title, journal, author, volume, date, issue and page fields.

The difficulty in parsing the 773/g (Host Item Entry/Related parts) field caused the frequency of volume, date, issue and pages to be slightly lower than the other fields. We believe that refining the extraction algorithm will help improve the recall of the data found in the 773/g. Expanding the extraction from one matched record to all of the records found in the matched Work cluster should also help improve the recall of fielded metadata. We also hypothesize that extracting over multiple records will help improve the quality/accuracy of the fielded metadata.

In addition to evaluating the total number of fields that were extracted we also had an interest in understanding how representative the extracted metadata was of the original citation. This statistic can be used to better understand how much structured data was mined from the matched record. In order to understand the coverage of extracted data, the resulting fielded metadata was compared to the original citation and a coverage percentage was calculated. The table below illustrates the results.
Although there were very few that had 90% - 100% coverage, it was very promising that nearly half of the test records had > 60% coverage. We think that making changes to the extraction code and including multiple records in the extraction process will help increase the coverage.

**Conclusion**

The process of converting a string citation into fielded metadata is a challenge that many institutional repositories face. Without fielded metadata, institutional repositories are not able to make their metadata visible to search engines (either traditional search engines such as Google or academic search engines such as Google Scholar). This study evaluated two methods for converting string citations into parsed metadata. Traditional citation parsing worked reasonably well but the time required to clean up the parsed data far exceeded the amount of time it would take to manually create the fielded metadata from a citation. This suggests that using open source parsing and manual review is not a scalable option for most institutional repositories. The second method for creating metadata was to match the citation against an OCLC MARC record and then extract fielded metadata from specific MARC tags. This method resulted in well-structured metadata. In order to better compare the two methods more comprehensive testing of the citation matching process will need to be conducted. After the algorithms for both the citation matching and metadata extraction are improved a second round of testing can be conducted and the results can be more accurately compared to those of the citation parsing method.

This process may soon be offered as a service for libraries that wish to structure their IR metadata for versatility and improved machine readability. Harvesting and indexing by Google Scholar is an immediate confirmed result of such structure.
Deliverable #4: SEO Improvements at University of Utah and MSU

On October 27, 2011 Google Scholar had indexed less than 1% (422 items) of the University of Utah’s 8,000+ scholarly papers housed in its open access IR, known as USpace. As of November 26, 2014 that indexing ratio had increased to approximately 48% (4,960 items) of the scholarly papers in USpace due to changes we implemented during work with the repository managers. Google Scholar has indexed ~2,160 additional digital collection items as scholarly works. For example, digital items from The Neuro-Ophthalmology Virtual Education Library (NOVEL) hosted by the University of Utah.

The result is that ~7,120 full text "scholarly papers" in PDF format are visible, accessible and free to the public. Without the SEO effort these papers would be more difficult to find or require a fee to access.

At Montana State University Library we implemented the recommendations from our paper "Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar" and have achieved nearly 100% indexing by Google Scholar (i.e., ~2,240 of 2,249 items).

• Other improvements at University of Utah
  o Increased Google Index Ratio among all digital collections from an average of 12% to 80%
  o Increased Google Index Ratio of USpace (Utah’s institutional repository) from 13% to 98%
  o Increased referrals from Google domains by 500%
  o Increased visitors to digital collections by 130%
  o Implemented scalable tools and repeatable processes
    ▪ Developed search engine friendly content sitemaps
    ▪ Institutionalized issue monitoring with Webmaster Tools
    ▪ Optimized server configurations for search engines
    ▪ Transformed IR metadata and reloaded
    ▪ Implemented Google Analytics to capture visitation traffic across all Utah library domains (previous approach was siloed).
    ▪ Influenced vendor product development
      • OCLC’s CONTENTdm digital asset management software
      • Ex Libris’ Primo discovery layer and Rosetta software

Deliverable #5: Influencing standards used by search engines

Our early research on IR indexing by Google and Google Scholar helped identify the
need to extend Schema.org for bibliographic citations. On September 13, 2011 the W3C established the Schema Bib Extend group "for extending Schema.org schemas for the improved representation of bibliographic information markup and sharing."2 The W3C Schema Bib Extend group was led by OCLC staff with whom we worked closely to help establish requirements. We also shared our efforts from Deliverable #3 (IR Data Modeling) to help influence solutions for extending Schema.org for bibliographic citations. Schema.org officially adopted the W3C Schema Bib Extend proposal for improving bibliographic citations and released Schema.org version 1.9 on August 8, 20143

**Ongoing Research**

While most of the deliverables described above developed from our research goals, we have uncovered additional areas of research that are incomplete but may lead to additional deliverables that other libraries will find useful.

**Web Analytics**

We have examined a sample of visitation and use statistics that libraries report for their websites, digital collections, and particularly, their institutional repositories. We have evidence that suggests libraries are both over-counting and undercounting visits to and downloads from their IR due to inappropriate configuration of web analytics software.

For example, by analyzing a five-day period of log file and Web analytics data from the USpace IR at the University of Utah we determined that Utah's Google Analytics did not count at least 125 unique Google Scholar users who downloaded at least 200 scholarly papers in PDF format. Given what we know about the cyclical usage of library digital assets, our best estimate is that USpace is failing to record between 8,000 and 11,000 PDF downloads annually – with the caveat that our estimates are based upon a very small data sample. These undercounted PDF downloads are only from Google Scholar visitors; other file types and visitors referred from other search engines must also be considered. We also verified all the major Web analytics services would not capture this high value metric, i.e. direct downloads of open access IR PDF files by Google Scholar.

Over-counting visits can be a problem, too, if libraries are using inappropriately configured log file analysis methods and are not filtering out known crawlers, spiders and scrapers, or the countless unknown usage that is clearly non-human behavior (e.g., requesting three, or more, web pages at the same time).

This phenomenon can lead to grossly inaccurate reporting to institutional administrators, funding organizations like IMLS, and governance or association bodies. We think we understand some of the reasons why this problem occurs and we believe there is significant additional research that needs to be conducted in this area by analyzing additional data sets from other libraries, by developing training.

---

2 [http://www.w3.org/community/schemabibex/](http://www.w3.org/community/schemabibex/)
3 [http://schema.org/docs/releases.html#v1.9](http://schema.org/docs/releases.html#v1.9)
programs, by developing standardized configurations for the implementation of web analytics software, and by publishing and presenting on this topic. With partners at OCLC Research, the Association of Research Libraries and the University of New Mexico we submitted a grant proposal to IMLS in late January 2014 and were awarded a new three-year National Leadership Grant to investigate this phenomenon more fully. The new grant officially kicked off on December 1, 2014.

**Semantic Identity**

In late 2012 a Google search for “Montana State University Library” produced a surprising result in Google’s Knowledge Card, the display that now commonly appears to the right of search results to provide immediate information about organizations and people. Instead of displaying the flagship MSU in Bozeman, MT, Google’s Knowledge Card in 2012 displayed a branch campus in Billings, MT. Further research revealed that the Bozeman library “property” had not been claimed in Google+, there was no article describing the library in Wikipedia, and the entry in Freebase was incomplete. All three of these situations have been remedied and the library now appears in its proper place in Google’s Knowledge Card, but this discovery opened yet another avenue for our SEO research.

Librarians have been late to embrace Wikipedia, and in fact have often actively discouraged engagement with what has become the world’s largest encyclopedia. But Wikipedia is much more than an encyclopedia of information for humans. Wikipedia establishes the legitimacy of entities and concepts for search engines, and a lack of presence in the online encyclopedia often means an organization simply don’t exist for search engines. While we have not yet conducted a systematic search of academic libraries, extensive spot-checking reveals that most libraries are either not represented at all in Google’s Knowledge Card, or the entry is not nearly as robust as it could be. This is an area where we believe significant research must be conducted, not only for better representation of library organizations, but also because many library concepts and services are poorly represented on the Semantic Web. This limits search engine comprehension of libraries and results in fewer user referrals.

**Impact of Structured Data Practices in Discovery and Use of Digital Collections**

Montana State University Library has started to implement HTML5 semantic tagging in our digital collections. Specifically, we are looking at how structured data practices (e.g., RDFa markup applying Schema.org vocabularies and linking to DBpedia Topics) create new understandings of digital collection content for software agents & machines. Three sample collections with the new markup are linked below.

http://arc.lib.montana.edu/schultz-0010/
http://arc.lib.montana.edu/msu-photos/
http://arc.lib.montana.edu/book/home-cooking-history-409/

A current thread of this research builds on the search engine optimization (SEO) work at Montana State University (MSU) Library and considers a control digital collection that has not been optimized (http://arc.lib.montana.edu/brook-0771/) versus a digital collection that has been built with semantic topics & machine-
actionable markup (http://arc.lib.montana.edu/schultz-0010/). To this end, we are also redesigning our optimized digital library application around three types of web pages as defined by Schema.org: about pages, collection pages, and item pages. Our community has an understanding of how to implement structured data; this research looks more closely at the question of why we should (or shouldn't) do it. We are starting to understand what can be gained by applying these structured data practices:

1. Allowing for machine-readable interpretations of our collections
2. Learning how to apply a web-scale classification system (Schema.org) and linked data topics to our collections
3. An ability to expose our collections data as an API or web service based on the structured data in the pages.
4. Discerning the impact that structured data has in creating sessions and page views of our collections through monitoring specific metrics within Google Analytics

In our preliminary findings, we are seeing spikes of engagement on our collection that has been optimized with structured data practices when compared with previous year's data and our “un-optimized” collection. See the figure below.

As this research continues, further testing and longitudinal analysis into the Return on Investment (ROI) of this RDFa and semantic HTML5 markup will need to be verified. The goal will be to derive best practices for SEO related to the markup and semantic tagging of digital collections.

Social Media Optimization Best Practices
The IMLS grant project included a component on Social Media Optimization (SMO), directed by the MSU Library Social Media Group (SMG). SMO is a set of practices and principles that aims to increase the shareability of Web content through online social networks with the overall goal of raising the awareness and usage of services and products. The practice of SMO is built on a foundation of social-focused metadata,
guided by the principle of enabling user-friendly sharing capabilities. Libraries can optimize web content for Twitter and Facebook, for instance, through the use of Twitter Cards and Facebook Open Graph tagging. Both offer the opportunity to provide descriptive information about Web content, which will then be included within the display of the Tweet or Facebook post. Twitter offers options to share images, audio, and video in the Twitter stream through Card tagging. As demonstrated in the images below, the information presented when a page has Twitter Cards is much more robust and eye-catching and, consequently, likely to be shared. The Twitter Card data surfaces a preview of the image, a title, an author, and a description as opposed to only the Tweet text and a link to the resource.

Figure 1: Tweet of page without Twitter cards
Likewise, Facebook Open Graph tags pull images, descriptions and titles of resources into a Facebook post. With more information immediately visible in these Facebook and Twitter posts, users are more likely to interact with the posted information and share it with their friends and followers. In addition to creating metadata that can be machine-harvested by major social media services, libraries can also offer user-facing social media share buttons on web pages to encourage and enable sharing across major platforms.
A comparison of the MSU Library's site before Twitter Cards and Facebook Open Graph tags were added (January 2013-October 2013) to the same time period one year later when SMO was applied (January 2014-October 2014) showed a jump in Facebook traffic to our optimized pages of 550 percent and Twitter traffic to these pages of 84 percent.

**Social Media Campaigns**

In order to understand more about shareability and cross-channel hashtagging, we developed several social media campaigns, explained below.

1. **#LibraryChamp** – The MSU Library SMG created a photobooth event during a major campus-wide festival, Catapalooza, just prior to the start of fall semester. The purpose was to draw students into this campaign while integrating the MSU Library’s presence, and promoting awareness of the library while also exploring targeted cross-channel hashtagging. Students were encouraged to have their photo taken with the MSU Mascot, Champ. These photos were subsequently posted via the library’s Twitter and Facebook accounts using the hashtag, #LibraryChamp. According to Facebook’s internal analytics tool, the reach of our #LibraryChamp posts ranged from 42 to 631. Posts that included photos performed better than those without. The #LibraryChamp campaign was our most successful both in terms of post and tweet engagement as well as in our overall approach to the campaign.

2. **Video Campaign** – SMG developed a video to better understand content sharing across various social media platforms. The video told the story of the library user experience within and outside of the confines of the library building. With a run time of 1 minute 30 seconds and a simple narrative, we attempted to create an engaging clip that is as shareable as it is informative. We used link tagging within Google Analytics to track views of the video and discovered that email is a highly effective avenue through which to share library content. We anticipated greater traction from Twitter and Facebook,
but numbers indicate that 56% of our audience was directed to our video from links shared via email. Email is not always included as a social media platform, but we’ve learned that on our campus email serves as a way to connect and share information across departments, colleges, and organizations. We also invested $5 in a promoted post to boost our video’s presence on Facebook. This resulted in only 25 sessions, whereas our non-promoted Facebook post resulted in 107 sessions. While the promoted Facebook post reached more users than the regular post did, this did not result in click-through and views.

3. On September 24, the library celebrated “Hug Your Library Day.” We took a photo of library lovers “hugging” the library building and posted it across social media platforms. This yearly, single-post event generated heavy engagement through shares. Those in the photo tagged themselves and shared it on their personal social media accounts, generated numerous likes. This particular event demonstrated the impact of posting photos of people. They opted in to sharing the library’s content with those in their personal social networks. With five shares and more than 1,500 people reached organically, this was by far our most popular Facebook post. We saw similar reach with our corresponding Twitter post with more than 2,500 people reached (called “Impressions” in Twitter Analytics) with 15 retweets, 9 favorites, and 3 replies which is our most popular Tweet to date.

Social Media Analytics Tools
A number of tools are available for analyzing social media activity. In selecting analytics tools, it is useful to consider what insights the library hopes to gain through the use of these tools. For example, some products may suggest that you follow certain accounts on Twitter because they are highly influential in the number of people they reach when posting. This same product might suggest that you unfollow a person because they have few followers and thus have less influence. There is not a one-size-fits-all product for libraries as the information sought will vary by institution, but having a clear idea of what the library hopes to learn through analytics tools will be a useful exercise prior to investigating options.

Two of the largest social media platforms used by libraries, Twitter and Facebook, have their own internal analytics, which are useful complements to third-party products. Twitter has an Application Programming Interface (API) that allows for querying and downloading of data for local analysis. In July 2014, Twitter released a new set of analysis tools (analytics.twitter.com), which conveys up-to-the-minute information. Beyond the typical reporting of retweets, favorites, and replies, Twitter Analytics offers helpful information, such as how many people viewed a given Tweet (i.e. impressions), how many engagements it received (e.g., click-throughs on links, viewing of photographs posted, etc.), and breakdowns of this data by the hour. Facebook also has an analytics component called Insights, which is built into Facebook Pages (pages of entities such as libraries, rather than personal pages of individuals). This area shows how many people viewed the posts, how many people liked and shared it, and how many post links were clicked. Both Twitter and
Facebook offer longitudinal considerations so that activity can be compared over time.

Third-party analytics tools offer additional perspectives about social media activity and the user community beyond native social media analytics tools. For this study, tools reviewed included SocialBro, ManageFlitter, BirdSong, and Commun.it. These tools were selected based on a literature and open web review of highly-ranked tools in the general social media community. Each of these tools includes basic analytics, information about those accounts following and being followed, and engagement. Some commercial products offer a free account and then more features at an additional cost. Commun.it (www.commun.it) is recommended because of its simplicity of use and suggested accounts to follow based on your existing activity. SocialBro (www.socialbro.com) is also recommended because of its detailed analytics. Both products have a modest tiered subscription cost structure. To accompany insights from these products, we also recommend using Google Analytics, free standard web analytics software currently used by many libraries. It includes social channel integration for viewing a range of social-related Web traffic, including social referrals and social-initiated user movements through a website.

**Communication Plan**

Our proposal included dissemination of the findings of our research through publications, presentations, and webinar training sessions. We have made significant contributions in each of these areas, and future publications are planned as our ongoing research is completed.

**Publications**


Presentations and Training


• Clark, Jason and Scott Young. “Metadata first: Using structured data markup and the Google Custom Search API to outsource your digital collections search index.” *Digital Library Federation Forum*, Austin, TX, November 4, 2013.


• Arlitsch, Kenning and Patrick OBrien. “SEO for Digital Repositories.”
  - Utah Library Association Annual Conf., Salt Lake City, UT, April 27, 2012
  - CNI Spring Membership Meeting, Baltimore, MD, April 2, 2012
  - OCLC TAI CHI Webinar, March 16, 2012 [http://www.youtube.com/watch?v=190D6QCk2ok](http://www.youtube.com/watch?v=190D6QCk2ok)
  - CONTENTdm Users Group, American Library Association Midwinter Conference, Dallas, TX, January 23, 2012
  - Western Archival Network (IMLS planning grant) meeting, University of New Mexico, Albuquerque, NM, January 12, 2012.
  - CNI Spring Forum, San Diego, CA, April 5, 2011