



Getting Found: Search Engine Optimization for Digital Repositories

Authors: Kenning Arlitsch & Patrick OBrien

This is the narrative of a funded proposal.

Getting Found: Search Engine Optimization for Digital Repositories

Assessment of Need

Print collections continue to diminish as a percentage of information resources, and libraries have spent millions creating digital repositories in an effort to remain relevant in the age of electronic publishing. Federal funds have provided considerable assistance in this process; the IMLS DCC Registry alone lists 389 collections that have received funding¹. Some repositories contain special collections materials and are aimed at the lay public, while institutional repositories focus more heavily on the academic community. Though most are designed to support the philosophy of free and open access, some, such as the catalogs of academic presses, are used in the conduct of electronic commerce. Despite differences in purpose and content, digital repositories hold enormous value and promise, by cost-effectively preserving cultural heritage and making unique materials accessible to a wide audience, and by supporting cutting-edge research workflows required for the conduct of e-science.

Digital collections are sometimes considered separately from institutional repositories, and from more commercially oriented efforts. But from the perspective of general-purpose search engines such as Yahoo or Google, they are more alike than different. All repositories consist of digital objects stored in databases; they run on web server technologies; they require descriptive metadata; and they are all subject to changes in search engine policies and practices. Repositories of all types will benefit from the work we are about to propose.

All of the digital repositories described above face a common challenge: being found by interested users in a crowded and ever-expanding sea of information available on the Internet. Getting found means they must be included in the indexes of major search engines. A 2005 survey by OCLC demonstrated that 89% of college students begin their research with Internet search engines, and that only 2% begin their research at the library.² Faculty search behavior is similar. A 2010 study of active faculty researchers at four major universities found that “researchers find Google and Google Scholar to be amazingly effective” for their information retrieval needs and accept the results as “good enough in many cases.”³ Digital repositories created by libraries will be invisible to users if their contents are not indexed in these search engines.

Unfortunately, many digital repositories are barely visible in the results from common search engines. In 2010 a team at the University of Utah conducted a randomized survey of 650 known objects across thirteen repositories comprising the Mountain West Digital Library (MWDL), and revealed a disturbing pattern: only 38% of digital objects searched by title were found in Google’s main index. Worse, this Google Search Engine Results Page (SERP) consisted mostly of links back to a search results screen in the local database, rather than linking directly to the objects. Only 15% of the hits on the SERP provided users with direct links to the objects.⁴ The known-item title searching method that we employed probably produced the best results

¹ Institute of Museum and Library Services Digital Collections and Content Registry. Retrieved January 13, 2011 from <http://imlsdcc.grainger.uiuc.edu/>

² De Rosa, Cathy, et al. “Perceptions of Libraries and Information Resources: A Report to the OCLC Membership”, OCLC, 2005. Question 520

³ Kroll, Susan and Rick Forsman. “A Slice of Research Life: Information Support for Research in the United States”, *OCLC Research and the RLG Partnership*, June 2010. Retrieved December 29, 2010 from <http://www.oclc.org/research/publications/library/2010/2010-15.pdf>

⁴ See summary of results of the MWDL repositories survey in Appendix A

possible at the time; searching by keyword or subject term would likely have presented even fewer items from the library's digital collections.

The Utah team reviewed the literature pertaining to SEO in libraries and found that most research was outdated and dealt with general websites. The minimal research dealing with digital repositories sometimes concluded by suggesting that libraries replicate content outside the database in a static format in order to make it friendlier to search engines, a method that seems arcane and burdensome. "Unless links are located on a static web page, crawlers won't find them, and many such links are not followed."⁵

In subsequent, informal searches of other random repositories across the country, we found results similar to the MWDL survey. Some repositories performed better, while others were worse. Some institutional repositories had good showings in Google, but not in Google Scholar, and most general repositories that we surveyed had almost no presence in Google Images. A presentation⁶ we gave at the *Digital Library Federation Forum* in November, as well as discussions with colleagues at other institutions helped us to confirm that this problem is widespread: most librarians are unaware how invisible their collections are to their users, and most institutions lack the skills necessary to understand and address the multiple layers required to improve search engine indexing and rank.

Why do repositories make such a poor showing in search engine results? The Utah research revealed a complex picture. We have identified several distinct problem categories that we intend to address:

1. Web servers may not be configured correctly to invite crawlers, and server speed performance may be unacceptably slow to search engines.
2. Repository software may present an impenetrable labyrinth for crawlers.
3. Metadata are often not unique or structured as recognizable taxonomies, and in some cases search engines may not desire the schema employed by the repository.
4. Search engine policies and practices change.
5. Generally accepted standards in the library community may not be honored or supported by search engines.

By addressing these categories in a pilot project at Utah, we demonstrated considerable success with selected collections on our CONTENTdm server. Our team significantly reduced crawler errors, developed a more efficient path for those crawlers to follow, and eliminated conflicts between Google sitemaps and server robots.txt files. We raised our average digital objects indexing rates in Google's main index from 12% in July 2010 to 46% in January 2011.⁷ One collection – the *Western Soundscape Archive* (WSA) – is a star; Google currently includes 93% of WSA objects in its index. On the other end of the spectrum is an Arabic Papyrus collection that registers only 11% in Google. The excellent metadata in the WSA, complete with taxonomy, may be partially responsible for its high indexing rate, but at the moment it's only a theory. There is a long way to go to configure solutions that will work well across different collections, institutions, and repositories, and we have only just begun to address metadata issues. We also have more

⁵ DeRidder, Jody L. "Googlizing a Digital Library", *Code4Lib Journal*, Issue 2, March 23, 2008

⁶ Arlitsch, Kenning, O'Brien, Patrick, and Sandra McIntyre. "Search Engine Optimization for Digital Collections", *Digital Library Federation Forum*, Fall 2010. Retrieved December 27, 2010 from <http://www.clir.org/dlf/forums/fall2010/25DLFseo.pdf>

⁷ See pilot project results in Appendix B

research to conduct on Google Scholar; the next section will show that its requirements and impact are different than Google's main index.

The solutions are not all technical in nature; they require coordination of different departments and types of personnel. Our project team is well-suited to develop useful solutions for SEO in digital repositories. The team has more than forty years of combined experience with the creation of digital repositories, management and administration, industry SEO, and metadata research and application. With OCLC as our partner, we intend to expand our research, develop a toolkit for digital repository managers, create automated mechanisms for transforming metadata, and offer SEO training for the library community. With support from the Mountain West Digital Library, we will be able to test solutions before offering them to a wider community, and with support from the Digital Library Federation we will effectively communicate our findings and solutions.

National Impact and Intended Results

As the economic recession has tightened university budgets, more emphasis is being placed on assessment and measurement of outputs. Recent research in the UK⁸ suggests that institutional repositories may play a crucial role in measuring research output, and in turn may affect university rankings. The *Times Higher Education* publishes an annual ranking of the top world universities, and research citations contribute 32.5% toward each university's score.⁹ The ability of institutional repositories to affect citation rates may be seriously hampered if their content is invisible to researchers who primarily use Google Scholar to begin their research.

Academic libraries have led the open access movement and the development of institutional repositories. But just as institutional repositories are gaining enough mass to make them useful and credible sources of research output, the difficulties associated with SEO threaten to undermine their potential. Our work addresses institutional repository needs by focusing on the special metadata requirements for Google Scholar.

Faculty and other authors who contribute publications to institutional repositories inevitably lose interest if they cannot locate their materials in Google Scholar. Donors of materials to library Special Collections and Archives are also important constituents, and if they cannot locate in Google's main index the digitized versions of materials they have gifted to libraries, or at least the finding aids describing them, they may be less likely to donate again. Part of our toolkit will assist with donor management and fundraising by capturing metrics that will show faculty and donors that their contributions can be found on the Web.

Librarians are great believers in standards, and while building digital repositories they have dutifully followed standards for scanning, metadata, harvesting, and web services, among others. Unfortunately, some search engines are not honoring some of those standards. The following two examples demonstrate the impotence of library standards in the search engine world:

⁸ Key Perspectives, Ltd (corporation). "A Comparative Review of Research Assessment Regimes in Five Countries and the Role of Libraries in the Research Assessment Process", *OCLC Research*, December 2009. Retrieved December 29, 2010 from <http://www.oclc.org/research/publications/library/2009/2009-09.pdf>

⁹ "World University Rankings 2010-2011", *Times Higher Education*. Retrieved January 5, 2011 from <http://www.timeshighereducation.co.uk/world-university-rankings/>

- In 2008 Google announced discontinuation of support¹⁰ for OAI-PMH, the metadata harvesting protocol that has been widely adopted in libraries.
- In August 2010, Google Scholar made the following announcement on its Webmaster Inclusion Guidelines site: “Use Dublin Core tags (e.g., DC.title) as a last resort - they work poorly for journal papers...”¹¹

Whereas repository managers may have correctly presumed they were being indexed more fully prior to the OAI-PMH change, many we spoke with recently were not aware of the announcement or the effect the change has had on their repositories' presence in Google's index. Our own experience at Utah points to a gradual purge of repository content from Google's indexes in the months that followed that change. The announcement about Dublin Core metadata was even more concerning, since most libraries' digital repositories use it to describe their collections. Instead of Dublin Core, Google Scholar recommends using schemas prevalent in the publishing industry, such as Highwire Press, Eprints, BE Press, and PRISM. Many librarians despair that they must now manually change their IR metadata schema to assure inclusion in Google Scholar. This proposal intends to assuage that burden on personnel and budgets by building automated transformation mechanisms.

Search engines can be thought of as “users with substantial constraints: they can't read text in images, can't interpret JavaScript or applets, and can't 'view' many other kinds of multimedia content”¹². Google and other search engines encourage adherence to W3C Web Content Accessibility Guidelines, a clearly defined standard to which libraries should aspire for ethical reasons in addition to SEO. Our proposal addresses specific means of applying accessibility standards to digital repositories.

Google claims the majority of direct search engine traffic, according to Comscore¹³, but the market is fluid, and other search engines, particularly Microsoft's Bing, are advancing. Bing powers the Yahoo! search in addition to its own site, and can claim nearly 30% of the search engine market in those combined sites.¹⁴ In recognition of the hunger for an understanding of SEO, Bing has launched official blogs for updates and “Bing Webmaster Tools, a toolset based on Google Webmaster Tools”.¹⁵ We will monitor the information on these sites, and use them to our advantage as we make recommendations for inclusion in Bing, as well as in Google.

Social media sites are making enormous strides as well and may change the face of Internet searching. Facebook's new Open Graph protocol brings an entirely new dimension, powering “a social search – one based on likes instead of links”¹⁶. In 2010, Facebook surpassed Google for

¹⁰ Mueller, John. “Retiring Support for OAI-PMH in Sitemaps” *Google Webmaster Central Blog*, August 23, 2008. Retrieved December 27, 2010 from <http://googlewebmastercentral.blogspot.com/2008/04/retiring-support-for-oai-pmh-in.html>

¹¹ “Google Scholar Inclusion Guidelines for Webmasters”. Retrieved December 27, 2010 from <http://scholar.google.com/intl/en/scholar/inclusion.html>

¹² Hagans, Andy. “High Accessibility is Effective Search Engine Optimization”, *A List Apart*, November 8, 2005. Retrieved January 8, 2011 from <http://www.alistapart.com/articles/accessibilityseo>

¹³ “comScore releases November 2010 U.S. search engine rankings”, *comScore Website Press Release*, December 15, 2010. Retrieved January 15, 2010 from

[http://www.comscore.com/Press_Events/Press_Releases/2010/12/comScore_Releases_November_2010_U.S._Search_Engine_Rankings/\(language\)/eng-US](http://www.comscore.com/Press_Events/Press_Releases/2010/12/comScore_Releases_November_2010_U.S._Search_Engine_Rankings/(language)/eng-US)

¹⁴ McGee, Matt. “Bing's market share up 51% in past 12 months”, *Search Engine Land*, August 25, 2010. Retrieved January 15, 2011 from <http://searchengineland.com/bing-market-share-up-51-percent-49368>

¹⁵ Pring, John. “SEO for Bing versus Google”, *Six Revisions*, October 23, 2010. Retrieved January 15, 2011 from <http://sixrevisions.com/content-strategy/seo-for-bing-versus-google/>

¹⁶ Ingram, Matthew. “Is Facebook's Social Search Engine a Google Killer?” *Gigaom*, June 25, 2010. Retrieved December 26, 2010 from <http://gigaom.com/2010/06/25/is-facebooks-social-search-engine-a-google-killer/>

total share of Internet traffic for the first time.¹⁷ Not one to be left behind, Google recently introduced a new indexing system known as Caffeine, which places more emphasis on content that is continually being refreshed and updated, like blogs and social media sites.¹⁸ As the content in library and archive digital repositories tends to be static, this new emphasis works against us. Our proposal intends to investigate and engage this paradigm shift as it occurs, and develop methods to address the social media emphasis.

The fluidity, the continuous innovation, and the intense competition in the search engine market make it a confusing and complex arena for the management of library operations. Librarians and other library staff generally lack the skills and knowledge to effectively develop and manage search engine optimization strategies. At best, repositories are harvested and indexed by search engines without librarian intervention or knowledge; at worst they are ignored. The search engine optimization world has a shady side, too, as “black hat” techniques are used by some web developers to game search engines and to rank their websites higher in SERP. Search engine developers are on constant guard against these techniques, in order to protect their users and to deliver the best experience to them. As a result they are often cagey about their harvesting and indexing requirements, and have been known to ban sites from their indexes for perceived infractions.¹⁹ It is important that libraries do not engage in black hat techniques, for any short-term gains will be outweighed by long-term failure. Our toolkit will provide guidance for repository managers, helping them to avoid pitfalls, and helping them to practice the “white hat” SEO techniques that are accepted and promoted by search engines.

The economic pressures mentioned at the beginning of this section are forcing libraries and archives to demonstrate their value proposition in quantitative terms. This proposal will help digital repositories meet that requirement. Repository managers will be able to monitor and report the numbers of digital objects that are being harvested and indexed by search engines. Correspondingly, success at that metric is expected to bring increased visitation and use of the repositories, demonstrating value for the institution and a return on investment for all funding sources.

Project Design and Evaluation Plan

Our project design builds on what we have learned from our own investigative efforts, carried on from April to November 2010. We are confident that our efforts will help libraries achieve better harvesting and indexing results from major search engines. Our goals and objectives are aligned along three general tracks:

1. Expand our 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.

¹⁷ McGee, Matt. “Facebook Passes Google (Again) as Most-Visited US Site: Hitwise”, *Search Engine Land*, March 15, 2010. Retrieved December 26, 2010 from <http://searchengineland.com/facebook-passes-google-again-as-most-visited-us-site-hitwise-38164>

¹⁸ “Our new search index: Caffeine” *The Official Google Blog*, June 8, 2010. Retrieved January 15, 2011 from <http://googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html>

¹⁹ Malaga, Ross A. “Worst Practices in Search Engine Optimization.” *Communications of the ACM*, Dec2008, Vol. 51 Issue 12, p147-150

Project Goals

1. **Goal: Gather additional data to demonstrate the current environment for digital repositories and search engine optimization**
 - a. Objective: Extend research to other search and social media engines
 - b. Objective: Expand and formalize survey to include repositories nationally
 - c. Objective: Gather and publish data to help the study of increased visibility of IR's in search engines on citation rates
2. **Goal: Improve repository hardware performance**
 - a. Objective: Establish benchmark performance indicators
 - b. Objective: Identify risk areas in web server configuration that will affect search engine optimization and server performance
 - c. Objective: Make recommendations for automated generation of sitemaps required by search engines
3. **Goal: Improve digital repository software for SEO purposes**
 - a. Objective: Develop recommendations for repository software developers
 - b. Objective: Create guidelines to help libraries make repository software selection decisions
 - c. Objective: Identify standards and methods that improve compliance with the Americans with Disabilities Act
4. **Goal: Improve metadata frameworks to make content more robust so that search engines can provide more relevant results to researchers.**
 - a. Objective: Develop crosswalks from Dublin Core to Highwire Press and other metadata schemas as required by search engines
 - b. Objective: Develop automated metadata transformation mechanism within OCLC's Digital Collections Gateway
 - c. Objective: Develop automated metadata transformation mechanism that can be adapted to local repositories
 - d. Objective: Conduct a discipline-specific case study of an ontology to make descriptive metadata more structured and relevant.
5. **Goal: Help digital repository managers measure and monitor their effectiveness in achieving visibility in search engines.**
 - a. Objective: Create effective reporting tool requirements for measuring key performance indicators used by library staff
 - b. Objective: Develop recommendations to manage expectations and increase participation from donors and author contributors
6. **Goal: Communication and Training (adoption of results)**
 - a. Objective: Provide training to the digital repository community, delivered through three webinars
 - b. Objective: Publish findings
 - c. Objective: Present findings at conferences

Timeline

Months 1-6: We will launch our grant work on November 1, 2011, and will first hire a half-time graduate student to begin researching the effects of other search and social media engines on digital repositories. We will submit a formal request to the David Eccles School of Business MBA field study program, with the aim of securing a team of MBA students for the spring semester. Their immediate task will be to help design and conduct an expanded and more sophisticated survey of repository performance in various search engines. We will gather data based not only

on known item searching, but also by keyword searching. Under the guidance of Arlitsch and O'Brien, the MBA team will also help establish a framework of metrics and begin gathering web log data that may eventually help others make the case for IR citation rates and university rankings. We will also begin to develop the toolkit, focusing first on hardware configuration recommendations.

Months 7-12: We will hire a second half-time graduate student who will help analyze data gathered in the expanded survey. Both students will help the SEO expert design reporting tools, based on the metrics framework devised by the MBA students. At OCLC, Godby will map crosswalks from Dublin Core to other metadata schemas recommended by Google Scholar, and merge them into OCLC's existing Common Data Framework. At the end of year one we will evaluate our progress and finalize the plan for year two.

Months 13-18: In the second year we will continue to develop the toolkit. The programmer from Utah will work with Godby and her team at OCLC to help develop metadata transformation mechanisms, both for OCLC's Digital Collections Gateway and as a solution that may be adapted to local repositories. Working with the graduate students, Arlitsch and O'Brien will begin to develop a training program. We will also begin to test toolkit solutions at participating MWDL sites.

Months 19-24: We will reach out to the major digital repository software developers (DSpace, OCLC's CONTENTdm, Fedora, Ex Libris' DigiTool, etc) with our recommendations for making the software friendlier to search engines. With Godby's guidance, we will investigate and test a discipline-specific ontology and vocabulary. At the end of year two we will evaluate our progress and finalize our plan for year three.

Months 25-30: In the third year we will launch our training program through a series of webinars, and begin writing with the aim of publishing our findings. Will finalize the toolkit based on feedback from the MWDL test bed sites, and we will step up the number of presentations at conferences to make others aware of our progress.

Months 31-36: Nearing the end of the grant period, we will conduct final training sessions, load the toolkit to the DLF website, and conduct a final evaluation of our progress. We will submit our final report to IMLS by November 30, 2014.

Methodology

To develop the toolkit we will use an iterative process that focuses on achieving the goals of increased visibility and use of digital repositories. Each phase will take approximately six months and will result in incremental improvements or additions to the frameworks, tools, and guidance provided in the toolkit. Each iterative phase is driven using Voice of the Customer techniques to define, prioritize, and align the goals, processes, and systems that contribute to the creation of digital repositories.

The toolkit will be presented in a generalized business process framework that digital repository stakeholders can recognize and understand. The framework will contain guidance and recommendations on the various business and system processes needed to set up and maintain a continuous improvement SEO program for digital repositories. This includes process goals, inputs (e.g., metadata schemas, relevant standards, key configuration files), outputs (e.g., stakeholder reports), process roles, key activities, and performance metrics. We will collect and analyze the data to design, optimize, and test alternative solutions using visual models and rapid

development prototyping. The toolkit will be based on principles of open source, allowing others to add and edit in the future as conditions change. We will publish the toolkit on SourceForge or a similar site under a Creative Commons “Attribution (CC BY) License”.²⁰

Evaluation Plan

Our proposal goals and objectives follow three general tracks: 1) gathering further data through surveys and investigative research; 2) developing a toolkit that will help repositories with SEO; and 3) disseminating our results through a training program, conference presentations, and publications. Some of our goals, such as the development of the toolkit, can be measured for success during the period of the grant, but the impacts of the knowledge and tools we plan to produce will only be realized several years after the grant is concluded.

To monitor progress toward project goals and assess impact, we will apply principles of Outcome Based Evaluation (OBE) and evaluate our success where we can. For instance, we will create surveys for participants in our training programs, where our target audiences will primarily be repository managers and their staff. We will be able to measure the effectiveness of our training by measuring whether participants felt they gained new knowledge and understand some of the changes they must make, or metrics they must monitor with regard to their repositories. Similarly, we will evaluate the success of our presentations by asking our audiences to fill out evaluation forms that indicate whether they have gained new knowledge and better understanding of SEO issues. Success will be measured against the Project Goals and Objectives we have stated above, but also by our communication and sustainability efforts.

Project Resources: Budget, Personnel, and Management

Budget

The total budget for this three-year project is \$878,820, of which we request direct costs of \$330,928 from IMLS. The University of Utah is contributing \$310,816 as cost share, and OCLC is contributing \$20,516, for a total cost-share of \$331,332. Indirect costs are calculated at the University of Utah’s federally negotiated service rate of 32.7% for a total of \$216,559.

The bulk of the budget is devoted to personnel. Patrick O’Brien will be hired as temporary University of Utah research (adjunct) faculty for the three years of the grant. We request funding for two half-time graduate students, one of whom will work for the duration of the project, while another will work for the second two years. In addition, we request funding to hire a team of MBA students from the University of Utah’s David Eccles School of Business field study program.

We also request funding for travel so that we may actively communicate our findings and recommendations at national conferences. Some travel funding will be required for on-site meetings with OCLC personnel.

Most of the work will be performed at the J. Willard Marriott Library, University of Utah. The library has already committed resources to investigating SEO issues with regard to its digital repositories by conducting survey and research work, by giving several presentations, by communicating with colleagues around the country, and by hiring Mr. O’Brien for one year.

Our formal partner on the grant is OCLC, Inc. They will contribute 5% of Jean Godby’s salary over the course of the three-year project. Dr. Godby will create the metadata crosswalks, and

²⁰ See Creative Commons licensing definitions. <http://creativecommons.org/licenses/>

will work with other OCLC staff and Utah's programmer to develop the transformation mechanisms.

Please see the Budget Justification form for a line-by-line explanation of each item.

Personnel

Principal Investigator **Kenning Arlitsch** is Associate Director for Information Technology Services and carries the rank of full librarian at the University of Utah's J. Willard Marriott Library. He launched Utah's digitization program in 2000, and is the founder of the *Mountain West Digital Library* and the *Utah Digital Newspapers* program, as well as co-founder of the *Western Soundscape Archive*. He and his staff are responsible for digitization, design and production of websites and interfaces, the library's ILS, its staff intranet, and the underlying server infrastructure. Mr. Arlitsch is a graduate of the *Frye Leadership Institute* and the *Research Libraries Leadership Fellows* program from the Association of Research Libraries, and has a proven track record with IMLS funding. He will oversee the project at 15% of his time.

Patrick O'Brien is an expert in customer focused, data driven sales and marketing operations. He specializes in the use of new media channels and Internet marketing to increase product visibility, acquire new customers, and improve customer satisfaction. He first began incorporating Search Engine Optimization (SEO) into demand generation marketing programs in 1997. He is a former Accenture Strategy Consultant with over 15 years experience working with business executives on converting marketing strategy into actionable results within the Pharmaceutical, Biotechnology, Healthcare, Financial Services and Telecommunications industries. Mr. O'Brien holds a B.A. in Economics from UCLA and a MBA in Marketing and Finance from The University of Chicago, Booth School of Business. Patrick will design and execute practical statistical models that measure collection key performance indicators for library staff. He will also perform statistical analysis to identify and measure impact of SEO methods, and he will create and publish data sets. He will devote 95% of this time to the grant.

Jean Godby, Ph.D. is a Research Scientist with OCLC who specializes in content analysis, subject classification, and the organization of full-text document collections. She is an internationally recognized expert, and has contributed significantly to the scholarly record on metadata. She currently manages the Metadata Schema Transformations project at OCLC. Dr. Godby will contribute 5% of her time to manage the development of a crosswalk from Dublin Core to the Highwire Press metadata schema, and will help develop automated transformation mechanisms.

Management

Mr. Arlitsch and Mr. O'Brien will guide the overall direction of the project. They will be responsible for planning, managing staff, disseminating findings, and evaluation; they are responsible for the toolkit products; and they will organize meetings or travel as necessary. They will also work closely with OCLC Research Scientist Jean Godby and her team to develop the crosswalks and metadata transformation mechanisms. Further, they will manage the efforts of the graduate students and the field study program (for one semester), as well as the programmer and web designer at the University of Utah.

The library will apply to the MBA field study program at the David Eccles School of Business at the University of Utah. The library has used this partnership once before for its ecommerce efforts, with good success. The field study program is designed to give approximately five MBA

students real-world experience by having them work as a team with an organization on a specific project for one semester. (See letter of support from Jeffery Brown)

Communication Plan

We intend to communicate and disseminate the findings of our research in three ways:

1. Presentations at relevant library conferences
2. Publication of data sets and papers
3. Webinar training sessions

The Digital Library Federation has offered its support as an informal partner. (See letter of support from Rachel Frick). DLF commits to helping us communicate our findings by facilitating the conversation and developing community activity around the SEO topic. The presentation at the DLF Forum last fall generated significant interest, and DLF will feature the topic again at its future forums to keep the community apprised of our work. In addition, DLF is developing a section of its website where community issues are discussed, through blogs and podcasts, and they will feature SEO topic there as well. The toolkit will be linked from this site to SourceForge or similar site, where it is most easily accessed and updated by other contributors according to the terms of the Creative Commons license.

The Mountain West Digital Library is another informal partner that has offered its support. The MWDL will serve as a test bed for our proposed solutions; this will give us an opportunity to test technical solutions on several different servers and repository software, in addition to testing administrative recommendations on different types and size of staff. The MWDL will also provide a perpetual link to the toolkit. (See letter of support from Sandra McIntyre)

Finally, our formal partner, OCLC, will assist in communicating our work through their venue at major conferences like ALA. OCLC may help us deliver our training sessions by providing webinar sessions and archiving them for future reference.

Sustainability

Search engine optimization for digital repositories cannot be addressed once and then left alone. As we have learned from our research thus far, it requires continually monitoring metrics at each repository for long-term success. The reporting tools we intend to develop and include in the toolkit will help repository managers monitor those metrics and adjust their practices accordingly to achieve success. Our presentations and publications will help make our work sustainable by passing on our findings and raising awareness in the community. The training we will provide will raise the level of expertise among digital repository managers across the nation.

The Digital Library Federation is an open community – anyone can join the conversation. Long an important part of the digital library community, DLF has recently been infused with new direction, energy, and a new home at the Council on Library Information Resources (CLIR). DLF has committed to help sustain the conversation about SEO for digital repositories by featuring our work at their Fall Forums. This will develop a community around the topic, and DLF can be an effective home for that community by providing a place to publish additional findings and modify the open toolkit as SEO evolves.