

Open Access Dissemination Challenges: A Case Study

Philip Young

University Libraries at Virginia Tech, Blacksburg, Virginia, USA

Abstract

Purpose- This paper explores dissemination, broadly considered, of an open access database as part of a librarian-faculty collaboration currently in progress.

Design/methodology/approach- Dissemination of an online database by librarians is broadly considered, including metadata optimization for multiple access points and user notification methods.

Findings- Librarians address open access dissemination challenges by investigating search engine optimization and seeking new opportunities for dissemination on the web. Differences in library metadata formats inhibit metadata optimization and need resolution.

Research limitations/implications- The collaboration is in progress and many of the ideas and conclusions listed have not been implemented.

Practical implications- Libraries should consider their role in scholarly publishing, develop workflows to enable it, and extend their efforts to the web.

Originality/value- This paper contributes to the scant literature on dissemination by libraries, and discusses dissemination challenges encountered by a non-peer reviewed, dynamic scholarly resource.

Keywords- Open access, Dissemination, Metadata optimization

Paper type- Case study

Introduction

The online environment has brought about a revolution in scholarly publishing. Scholars now publish on the web, whether in an online journal or to their own web site. They are also creating new forms of scholarship and providing access to scholarly resources. This paper discusses a database that came online in 2006, “Spenser and the Tradition: English Poetry 1579-1830” [<http://spenserians.cath.vt.edu>]. The database was brought to the attention of librarians at Virginia Tech by its creator, Dr. David Radcliffe, a faculty member in the English department. The database, more than 15 years in the making and still being updated, offers full text of poems, although its uniqueness lies in its links between poets as authors and readers—in short, a genealogy of influence. The complex web of links between poets is far easier to express and explore in the digital environment. The database is written in MySQL and is hosted on a server by the Center for Technology in the Humanities [<http://wiz.cath.vt.edu/cath/cath.html>].

Although the database is not hosted by the University Libraries at Virginia Tech, Dr. Radcliffe sought help from librarians for its dissemination—clearly, just putting it online wasn’t enough. Faculty are turning to librarians for assistance in producing electronic resources (Brown et al. 2007) and the trend is likely to continue. Dr. Radcliffe’s interests

include the cataloging of the database so it can be entered into Virginia Tech's library catalog (as well as other library catalogs), and how to ensure that it ranks highly in the results returned by search engines. What information should he provide on his website, or in his code, to best enable bibliographic and web indexing?

This faculty-library collaboration faces several challenges. First, libraries are not often called upon to assist in the dissemination of information in the broadest sense. In most cases, simply making a scholarly resource available online qualifies as dissemination. However, this collaboration chose to examine dissemination more deeply, in terms of how metadata might be optimized for a variety of online dissemination purposes, and more broadly, in terms of functions traditionally associated with publishers. Relatively little has been published about optimizing metadata (Dawson 2005). Second, this case study concerns a scholarly resource rather than scholarship itself. As such, it is not peer-reviewed, and its dynamic, updating format offers challenges.

While open access (OA) definitions can be extensive, for the purposes of this paper OA simply means "digital, online, free of charge, and free of most copyright and licensing restrictions" (Suber 2007). A search reveals little published literature on the dissemination of OA resources. Dissemination has been identified as the most difficult problem of the OA movement (Morgan 2004). A broad exploration of dissemination possibilities for the English Poetry database suggests that librarians could optimize metadata for cataloging, repository harvesting, and the web, as well as find ways to notify interested users of the resource.

Dissemination was an integral role of the earliest libraries, but over time publishing and libraries became separate (Grafton 2007). The two are converging again in the digital environment. A recent report urges universities to renew their commitment to publishing by combining skills at libraries, university presses, and information technology departments (Brown et al. 2007). Libraries should seek new ways to support scholars in the online environment (ACRL 2007, Morgan 2004, OCLC 2004), and should focus on services including "the provision of a mechanism for the dissemination of information" (Manuel and Oppenheim 2007).

Cataloging

The provision of access best known by librarians is traditional cataloging using machine-readable cataloging (MARC) records. Online resources pose unique challenges to catalogers. Important information is not found in a standard place, and is often missing altogether. In recent years, several methods to extract metadata automatically from a web site's source code have been created (Library of Congress 2005, OCLC 2007, Su et al. 2002). However, in most cases catalogers will need information from the plain text of the site as well as the source code. Catalogers will typically look for a clear and consistent title, a statement of responsibility, a summary or description, software requirements (if any), an indication of whether the resource is updated (and with what frequency), and a date or dates, among other information. Sometimes this information

can be found on a web site's "About" page, if present. However, the web site's source code header will likely become the de facto provider of metadata, since it provides a place for structured metadata that can be extracted automatically.

The use of OCLC's metadata extraction tool (OCLC 2007) for the cataloging of the English Poetry database proved unsatisfactory, despite the tool's crosswalk for more than 1,000 metadata fields (OCLC 2007a). The poor result, however, is likely due to the lack of metadata in the source code (the "meta" tags were limited to author and keywords) rather than the performance of the tool. Catalogers rarely have influence on the coding of the websites that they are cataloging, but in such situations, how should a website be coded to obtain maximum value from the metadata extraction tool? According to OCLC (2007a), the tool will work better as more metadata is provided.

Metadata harvesting to create MARC records is unlikely to result in complete records, but can limit cataloger intervention to a minimum. Automated, or at least hybrid, means of metadata creation are needed (Weibel 2005), and are already in place at some abstracting and indexing services, as well as newspapers like the New York Times (Harris 2007). Additional cataloger time can be saved by using constant data in combination with metadata extraction. The proliferation of online updating resources like the English Poetry database calls attention to the need for automated updating of MARC records. Metadata extraction tools should be designed to check updating resources on a regular basis. Repository aggregators harvest metadata regularly, and Google recommends that webmasters use an "if-modified-since HTTP header" so its crawlers will recognize updated content (Google 2007). E-mail alerts for updated content have been employed by the Library of Congress (2005), although this works better for discrete rather than integrated updates.

Once the full catalog record is created, it can be entered into the OCLC database and exported to the local catalog. Once in OCLC's database, the record appears in WorldCat [<http://www.worldcat.org>]. The resource record includes a list of libraries that have added it to their catalog. WorldCat can be accessed and searched by anyone, and more importantly, its contents are indexed by search engines through Open WorldCat [<http://www.oclc.org/worldcat/help/en/#howget>]. Likewise, the record can be found in library catalogs that are pilot testing WorldCat Local, an implementation of the OCLC database as a local catalog [<http://www.oclc.org/news/releases/200659.htm>].

Repositories

Although the English Poetry database is on the web, there are advantages to also depositing the database in a repository. For example, a repository would address digital preservation, especially format migration (Davis and Connolly 2007, SHERPA 2006). This aspect of repositories is especially important to Dr. Radcliffe, who has already experienced two format migration problems in the long gestation of the database. Repositories, and especially aggregation interfaces like OAIster [<http://www.oaister.org>], provide another means of user discovery, and better ranking in search engine results

(SHERPA 2006). Additionally, automated processes offer the potential to create a record for each of the 25,000 items in the database rather than one record for the database as a whole. Access to authors too numerous to list in a MARC record would be greatly improved.

However, repository deposit presents several problems for the English Poetry database. First, some repositories define open access in such a way that it would exclude the database either by format or its lack of peer review. These repositories are restricted to works of scholarship rather than scholarly resources, although the Open Access Initiative (OAI) was originally intended for non-peer reviewed materials (Rusch-Feja 2002) and most repositories use a broader definition of open access (Hood 2007). Second, repositories use a download model (Smith et al. 2003) that is not as user-friendly for a resource like a database, which is easier to use on the web. Third, updates to the database are problematic because repositories are designed to save versions of resources. This seems incompatible with an environment in which scholars increasingly will be contributing growing data sets (ACRL 2007). Depending on the frequency of the updates, the repository version could be out of synch with the web version, and there would be no need to save numerous older versions. Preservation and access functions conflict, in a similar way that remote storage for physical items enhances preservation while hampering access. Differences between repository software implementations in the handling of these resources have implications for repository selection and policy. Fourth, an appropriate repository may not be available for a particular resource. Dr. Radcliffe has been unable to identify a disciplinary repository for the database, and Virginia Tech does not yet have an institutional repository. Lack of repository access has implications for numerous subject areas, particularly in the humanities, which have been slower to develop them than in the sciences (Suber 2004), and for independent and developing world scholars. A third type of repository, the static repository, offers a low-effort, low-cost option (Moffat 2006), but is not well-suited for the size and dynamic nature of the database. A new initiative called Object Reuse and Exchange may provide a better model for complex objects like databases, and improve search engine behavior (OAI 2007).

The Open Access Initiative-Protocol for Metadata Harvesting (OAI-PMH) requires metadata exposed as simple Dublin Core, but encourages additional metadata formats, including MARC (Moffat 2006, OAIster 2007). Dublin Core metadata would also assist OCLC metadata extraction, although its simplicity would limit its usefulness in the more detailed MARC record. And even Dublin Core's minimal metadata scheme is not fully utilized by current OAI-PMH data providers (Ward 2004). The significant discrepancies in metadata detail between repositories and library catalogs must be addressed to achieve metadata optimization.

Search engine optimization

“Despite a mantra of interoperability, attention is rarely given to the question of how to ensure that meticulously crafted metadata is used beyond the confines of

its immediate surroundings. The existence of search engines is ignored or denigrated.” (Dawson and Hamilton 2006)

Search engines are the primary means of discovering and selecting digital content (Dawson 2004). The prominence of Google and other search engines among searchers has been noted, as well as increased collaboration between information providers and Google (Tenopir 2004). A study has shown that 89% of college students use search engines to begin an information search (De Rosa et al. 2006), and some claim that 95% of scholarly inquiries start at Google (Grafton 2007). Even in the hard sciences, search engines are a common choice. Kahn and Drey (2002) found that Google was the second choice of analytic and organic chemists, and the first choice among chemists in management and development positions. Search engines are easily accessible, and that is the most important variable governing the use of information (Morville 2005).

The practice of coding websites for the highest possible search result ranking is referred to as search engine optimization (SEO), and its importance for scholarly resources cannot be overstated:

To reach users wherever they are, we as a community need to disclose more metadata to OAI harvesters [and] Web crawlers... search engine optimization is crucial. (Smith-Yoshimura 2007)

Connecting users with the content and services we design and build is part of our broader mission. It's not good enough to create a great product and expect someone else to worry about how people will find it. Together with form and function, findability is a required element of good design and engineering. I relentlessly make this case to government agencies and nonprofits that don't have marketing departments. They tend to shy away from SEO as overly commercial, but they're missing a great opportunity to fulfill their mission by helping people find what they need. (Morville 2005)

OA-OAI archiving and Google indexing are completely compatible. We can do both, and we should. (Suber 2004a)

Designing scholarly web resources for high placement in search results makes sense. In addition to increased visibility, top results are perceived as authoritative (Morville 2005). When the disciplinary repository ArXiv [<http://arxiv.org>] redesigned its site for improved indexing by Google, usage increased 50% (Inger 2004).

Search engine optimization has largely been employed by commercial web sites. Because Google and other search engines do not reveal the details of their search algorithms, a small industry has been created to help webmasters optimize the coding of their web site so the site appears as high as possible in search engine results. Search engines also offer advice to webmasters for optimal indexing (Google 2007). Much of SEO is geared toward making sites easy to access and navigate by the crawlers, or automated robots, that are used by search engines to index the web. Generally, positive

factors for indexing include clear title tags (Dawson 2004, Sullivan 2002) and alternate text for images tags (Google 2007), a site map, incoming links (Brooks 2004), and top domain (Dawson and Hamilton 2006). Negative factors are primarily those inhibiting the crawlers, such as frames, JavaScript (Weideman and Schwenke 2006), Flash, and redirects (these factors also inhibit the metadata extraction tools used in cataloging). Some of these features increase usability yet are in conflict with SEO (Bosworth 2007, Weideman and Schwenke 2006).

Some aspects of the English Poetry database's source code, such as its extensive use of JavaScript, are in this category. Some steps to improve the indexing of the database have already been taken. Frames have been eliminated, alternate text for images added, and the original "org" domain reverted to "edu." Dawson and Hamilton (2006) advise that Google seems to privilege "gov", "edu" and "ac.uk" domains, so one should avoid using other domains merely to give a special project a memorable URL. Specific title tags (Dawson 2004) for each record in the database are in place. The influence of JavaScript on indexing can be mitigated, and a site map added. The database needs usability improvements of the kind that should not affect indexing. Preliminary feedback from reference librarians indicates that undergraduates will have difficulty navigating the database, and will expect a search box. The database currently enables searching, but not on the front page. In addition to a built-in search box, OpenSearch [<http://www.opensearch.org/Home>] code can be added to allow toolbar searching of the database, so that users can easily search the database wherever they are on the web. As a scholarly resource, citations for each record in the database should be provided. A stable URL is an important part of the citation. Permanent links can appear on each page, and full citations could be generated automatically from Dublin Core or other metadata (Jorgensen 2005), or exported to citation software.

SEO differs significantly from the cataloging and repository worlds, where explicit metadata is highly valued. Metadata extraction, for example, performs better with more metadata (OCLC 2007a). However, search engines mostly ignore metadata added by webmasters due to a history of abuse and misrepresentation (Beall 2006, Brooks 2004), particularly keywords (Dornfest et al. 2006, Sullivan 2002). While the importance of incoming links is frequently cited (Brooks 2004), Dawson and Hamilton (2006) demonstrate that a library resource can achieve top listing without any incoming links.

Other dissemination methods

While metadata optimization can enhance access, more explicit methods of dissemination deserve examination. Some of these methods are currently employed by publishers to notify interested users, and others emerge from the increasing interactivity of the web. This kind of dissemination can result in the incoming links that further enhance search engine indexing. Submission of a site's URL to search engines, directories and portals is one method. Search engines recommend site submission as part of their guidelines for webmasters (Google 2007). Indexing by Google Scholar was pursued for the database since it is a scholarly resource, but the search engine is limited to scholarship, that is,

textual narrative in the form of articles and books, much in the same way that some repositories are restricted.

A number of general portals encourage submission, such as the Open Directory Project [<http://www.dmoz.org>], the Yahoo Directory [<http://dir.yahoo.com>], and Intute [<http://www.intute.ac.uk>]. Mattison (2006) provides an extensive overview of disciplinary portals in the humanities. Among the portals linking to the database are two well-known sites, Voice of the Shuttle [<http://vos.ucsb.edu>] and Early Modern Resources [<http://earlymodernweb.org.uk/emr>]. However, Dr. Radcliffe reports a portal submission success rate of only 1 in 4, which was discouraging enough to make him give up. Sowards (1999) likewise found little success with URL submission to portals until news events created interest in his content. Many journals are now online and publish reviews of scholarly resources. An online review provides awareness as well as an incoming link.

The increasing interactivity of the web offers opportunities for dissemination. Lally and Dunford (2007) relate the use of the online encyclopedia Wikipedia to drive usage. In most cases this simply involves adding a link to a relevant article, although it sometimes entails writing a new article. An examination of incoming links to the database found several links from Wikipedia already in place. This examination also revealed that a link from a community blog such as MetaFilter [<http://www.metafilter.com>] or MonkeyFilter [<http://monkeyfilter.com>] can greatly increase awareness. Disciplinary mailing lists notify recipients of new resources, and the database received mention on the Byron list. Librarians can use collection development lists to alert other libraries that might want to add the database to their catalog. Hood (2007) suggests adding new online resources to pathfinders and subject guides, and the use of targeted e-mail alerts by subject bibliographers. Really Simple Syndication (RSS) has potential, but may not be appropriate where resources are changing frequently or numerous resources come online at once (as in the case of repositories). Publishers commonly use RSS table of contents alerts, print flyers, mail postcards, advertise in journals, and send e-mail. While publishers have more tools for generating awareness, their content is hidden behind a subscription wall. The general public and libraries that cannot afford a subscription have no access. Ironically, it is the OA resources like the English Poetry database which are difficult to disseminate, and librarians should be creative as possible in assisting their faculty in doing so.

Concluding discussion

The future of this collaboration in dissemination involves numerous tasks. The source code of the database's web site needs refinement following the recommendations of Brooks (2004) and especially Dawson and Hamilton (2006). A full metadata header and citation functionality need to be added, and navigation and search tools improved. Then more explicit dissemination methods can be employed. In addition, Dr. Radcliffe would like to produce a guide for other faculty who are interested in the dissemination of their online content. Libraries may want to consider similar recommendations as a service to their faculty, particularly since the proliferation of digital centers on campuses means that

much online scholarship will not be produced or hosted by libraries. Also, digital centers may not be as familiar with metadata uses for multiple purposes.

Measurement of access and dissemination after applying SEO principles will be a difficult task due to the variety and simultaneous influence of factors. Measurement might include indicators such as rank in search engines, library holdings in WorldCat, direct linking by libraries, statistics from website management applications as well as from the server and MySQL database, number of incoming links, success rate with portal submission, citations in scholarly papers, and improvements in metadata extraction.

Metadata optimization is necessary if efficiency and effective dissemination are to be realized by libraries. The current standard for online resources is some combination of extensible markup language (XML), Resource Description Framework (RDF), and Dublin Core (DC). Repository harvesting requires DC metadata and XML as its syntax; DC can be used in metadata extraction for cataloging; and RDF will be necessary for any future Semantic Web implementation. Also, citations can be extracted from DC. Libraries must first bridge the gap between Dublin Core and MARC. Omitting digital collections from the catalog results in information “silos” and requires users to search in different places. Workflows for electronic resources (such as electronic theses and dissertations) have already been created in libraries, and a similar but more comprehensive workflow should be created that provides access in catalogs, repositories, and on the web. One workflow integration tool recently became available (OCLC 2007b) that addresses the metadata problem by starting with the MARC record and deriving qualified DC upon deposit of the digital resource. This crosswalk direction may prove more effective than deriving MARC from DC.

Libraries must address the problem of providing access to online resources in an environment that largely ignores explicit metadata. The XML/RDF/DC metadata scheme may be useless to most search engine crawlers, yet the web is where most information seekers are going first. The invisibility of metadata to search engines may be one cause of so little effort by libraries toward SEO (Beall 2006). While this metadata scheme is compatible with the Semantic Web, much skepticism remains about user-supplied metadata (Weibel 2005). A more realistic scenario in which metadata could be fully utilized is that of “closed applications” (Brooks 2004) such as intranets or digital libraries, or dividing the web by top domain or other means. Google Scholar’s harvesting of citations from scholarly publishers may be one example. Until ways can be found to utilize the XML/RDF/DC scheme in web indexing, libraries should probably heed the recommendations of Dawson and Hamilton (2006).

Genre will become increasingly important in the online environment (Morville 2005). The library community may want to consider metadata identification of databases and other online resources, as well as creating a genre for scholarly resources. A category for scholarly resources (i.e., the materials on which scholarship is based) may become important as more primary research material is digitized and as more data sets are made available.

While faculty-created online resources such as the English Poetry database may not be common, or faculty not as concerned with dissemination, libraries should consider their role in scholarly publishing. As Manuel and Oppenheim (2007) state, “Google, repositories and libraries all have a part to play in improving dissemination, and thus research impact.” The knowledge of metadata and open access in libraries positions them well for increased faculty collaboration. As the volume of information increases, our ability to find particular items decreases, and we spend more time searching (Morville 2005). Online resources need more metadata, and libraries can fill the need in the scholarly arena.

References

- ACRL (2007), “Establishing a research agenda for scholarly communication: a call for community engagement”, available at: <http://www.ala.org/ala/acrl/acrlissues/scholarlycomm/SCResearchAgenda.pdf>
- Beall, J. (2006), “The death of metadata”, *The Serials Librarian*, Vol. 51, No. 2.
- Bosworth, A. (2007), “Google is destroying the web and you don’t even know it”, available at: http://blog.alexbosworth.net/article/google_destroying_the_web
- Brooks, T.A. (2004), “The nature of meaning in the age of Google”, *Information Research*, Vol. 9 No. 3, paper 180, available at: <http://InformationR.net/ir/9-3/paper180.html>
- Brown, L., Griffiths, R., and Rascoff, M. (2007), “University publishing in a digital age”, *Ithaka Report*, available at: <http://www.ithaka.org/strategic-services/Ithaka%20University%20Publishing%20Report.pdf>
- Davis, P.M. and Connolly, M.J.L. (2007) “Institutional repositories: evaluating the reasons for non-use of Cornell University’s installation of DSpace”, *D-Lib Magazine*, Vol. 13 No. 3/4, available at: <http://www.dlib.org/dlib/march07/davis/03davis.html>
- Dawson, A. (2004), “Creating metadata that works for digital libraries and Google”, *Library Review*, Vol. 53, No. 27, pp. 347-350, available at: <http://eprints.cdlr.strath.ac.uk/2289/01/ad200402.htm>
- Dawson, A. (2005), “Optimising publications for Google users”, pp. 177-194, in Miller, W. and Pellen, R.M. (eds.), *Libraries and Google*, Haworth Press, Binghamton, N.Y.
- Dawson, A. and Hamilton, V. (2006), “Optimising metadata to make high-value content more accessible to Google users”, *Journal of Documentation*, Vol. 62 No. 3, pp. 307-327

De Rosa, C. et al., (2006), College Students' Perceptions of Libraries and Information Resources, OCLC, Dublin, Ohio.

Dornfest, R., Bausch, P. and Calishain, T. (2006), Google Hacks (3rd ed.), O'Reilly, Sebastopol, Calif.

Google (2007), "Webmaster guidelines", available at:

<http://www.google.com/support/webmasters/bin/answer.py?answer=35769>

Grafton, A. (2007), "Future reading: digitization and its discontents", *The New Yorker*, available at: http://www.newyorker.com/reporting/2007/11/05/071105fa_fact_grafton

Harris, J. (2007), "Messing around with metadata", *Open*, [New York Times blog], available at: <http://open.blogs.nytimes.com/2007/10/23/messing-around-with-metadata/>

Hood, A.K. (2007), "Open Access Resources: Executive Summary", SPEC Kit 300, pp. 11-14, Association of Research Libraries, Washington, D.C., available at: <http://www.arl.org/bm~doc/spec300web.pdf>

Hunter, P. and Guy, M. (2004), "Metadata for harvesting: the Open Archives Initiative, and how to find things on the Web", *Electronic Library*, Vol. 22 No. 2, pp. 168-174, available at: <http://homes.ukoln.ac.uk/~lispjh/tel-metadata/metadata-final5.pdf>

Inger, S. (2004), "Google vs traditional information services: a comparison of search results", National Federation of Abstracting and Indexing Services (NFAIS), 22 February, available at: <http://www.scholinfo.com/GoogleversusTraitionalInformationServices.pdf>

Jorgensen, P. (2005), "Citations in hypermedia: implementation issues", *Information Technology and Libraries*, Vol. 24 No. 4, available at: <http://news.ala.org/ala/lita/litapublications/ital/volume242005/number4december/contentv424/jorgensen.pdf>

Kahn, D. and Drey, J. (2002), "Finding Chemical Information on the Web - the User's Viewpoint", *Free Pint*, Issue 109, available at: <http://www.freepint.com/issues/040402.htm#feature>

Lally, A.M. and Dunford, C.E. (2007), "Using Wikipedia to extend digital collections", *D-Lib Magazine*, Vol. 13 No. 5/6, available at: <http://www.dlib.org/dlib/may07/lally/05lally.html>

Library of Congress (2005), "Web Cataloging Assistant", Bibliographic Enrichment Advisory Team, available at: <http://catdir.loc.gov/catdir/beat/webcat.html>

Manuel, S. and Oppenheim, C. (2007), "Googlepository and the university library", *Ariadne*, Issue 53, available at <http://www.ariadne.ac.uk/issue53/manuel-oppenheim/>

- Mattison, D. (2006), "The digital humanities revolution", *Searcher*, Vol. 14, No. 5, pp. 25-34.
- Moffat, M. (2006), "Marketing with metadata—how metadata can increase exposure and visibility of online content", *New Review of Information Networking*, Vol. 12, Nos. 1-2, pp. 23-40.
- Morgan, E.L. (2004), "Open access publishing", available at <http://infomotions.com/musings/open-access/open-access.pdf>
- Morville, P. (2005), *Ambient findability*, O'Reilly, Sebastapol, Calif.
- OAI (2007), "Open Archives Initiative announces public meeting on March 3, 2008 to release Object Reuse and Exchange specifications", press release available at: <http://www.openarchives.org/ore/documents/ore-hopkins-press-release.pdf>
- OAister (2007), "How to become a data contributor", available at: <http://www.oaister.org/dataproviders.html>
- OCLC (2004), *The 2003 OCLC Environmental Scan: Pattern recognition*, OCLC, Dublin, Ohio.
- OCLC (2007), "Cataloging: Create Bibliographic Records", OCLC Connexion Client Guides, pp. 15-23, available at <http://www.oclc.org/support/documentation/connexion/client/cataloging/createbib/createbib.pdf>
- OCLC (2007a), E-mail from OCLC Connexion-Support, August 20, 2007.
- OCLC (2007b), "Attach digital content to WorldCat records", available at: http://www.oclc.org/support/documentation/connexion/client/cataloging/bibactions/#cat_act_attach_digital_files
- Rieh, S. Y., Markey, K., St. Jean, B., Yakel, E., and Kim, J. (2007), "Census of institutional repositories in the U.S.: a comparison across institutions at different stages of IR development", *D-Lib Magazine*, Vol. 13 No. 11/12, available at: <http://www.dlib.org/dlib/november07/rieh/11rieh.html>
- Rusch-Feja, D. (2002), "The Open Archives Initiative and the OAI Protocol for Metadata Harvesting: rapidly forming a new tier in the scholarly communication infrastructure", *Learned Publishing*, Vol. 15, No. 3, pp. 179-186.
- SHERPA (2006), "Fifteen common concerns- and clarifications", available at: <http://www.sherpa.ac.uk/documents/15concerns.html>

Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., Stuve, D., Tansley, R., and Walker, J.H. (2003), "DSpace: an open source dynamic digital repository", *D-Lib Magazine*, Vol. 9 No. 1, available at:

<http://www.dlib.org/dlib/january03/smith/01smith.html>

Smith-Yoshimura, K. (2007), "RLG Programs Descriptive Metadata Practices Survey Results", OCLC Programs and Research, available at:

<http://www.oclc.org/programs/publications/reports/2007-03.pdf>

Sowards, S.W. (1999), "Practical lessons for small-scale web publishers", *Journal of Electronic Publishing*, Vol. 5, Issue 2, available at: <http://www.press.umich.edu/jep/05-02/sowards.html>

Su, S.T., Long, Y. and Cromwell, D.E., (2002), "Metadata by crawling E-publications", *Information Technology and Libraries*, Vol. 21 No. 4, available at:

<http://news.ala.org/ala/lita/litapublications/ital/2104su.cfm>

Suber, P. (2004), "Open access in the humanities", *SPARC Open Access Newsletter*, Issue 70, available at <http://www.earlham.edu/~peters/fos/newsletter/02-02-04.htm#humanities>

Suber, P. (2004a), "The case for OAI in the age of Google", *SPARC Open Access Newsletter*, Issue 73, available at <http://www.earlham.edu/~peters/fos/newsletter/05-03-04.htm#oai-google>

Suber, P. (2007), "Open Access overview", available at:

<http://www.earlham.edu/~peters/fos/overview.htm>

Sullivan, D. (2002), "Death of a meta tag", available at

<http://searchenginewatch.com/2165061/print>

Tenopir, C. (2004), "Is Google the competition?", *Library Journal*, Issue 6, available at: <http://www.libraryjournal.com/article/CA405423.html?display=Online+>

Ward, J. (2004), "Unqualified Dublin Core usage in OAI-PMH data providers", *OCLC Systems & Services*, Vol. 20, No. 1, pp. 40-47.

Weibel, S.L. (2005), "Border crossings: reflections on a decade of metadata consensus building", *D-Lib Magazine*, Vol. 11 No. 7/8, available at:

<http://www.dlib.org/dlib/july05/weibel/07weibel.html>

Weideman, M. and Schwenke, F. (2006), "The influence that JavaScript has on the visibility of a website to search engines- a pilot study", *Information Research*, Vol. 11 No. 4, available at <http://informationr.net/ir/11-4/paper268.html>

About the author

Philip Young has served since 2006 as Catalog Librarian for science and technology at the University Libraries at Virginia Tech. Philip Young can be contacted at: pyoung1@vt.edu.