June 2012

SAGE White Paper on Discoverability in the Twenty-First Century

Mary M. Somerville
University of Colorado, Denver, Mary.Somerville@ucdenver.edu

SAGE's White Paper on Discoverability in the Twenty-First Century: Collaboration Opportunities for Publishers, Vendors, and Librarians

Fiesole Collection Development Retreat, "Information Discovery: Examining Enabling Technologies and New Tools," a Pre-Conference Presentation, April 12, 2012, Fiesole, Italy

by **Mary M. Somerville, MLS, MA, PhD** (University Librarian and Professor, University of Colorado Denver, and Director, Auraria Library, Denver, CO, USA) <mary.somerville@ucdenver.edu>

Abstract

Discoverability enables scholars to locate the content needed to advance their research and other creative activities. Improved discovery experiences require heightened collaboration among (1) scholarly publishers and their published authors; (2) search engine developers, database providers, abstracting and indexing services, and academic publishers; (3) electronic resource management and integrated library system vendors; and (4) librarians who advance institutional discoverability. To further exploratory cross-sector conversations, **SAGE** commissioned a White Paper, "*Improving Discoverability of Scholarly Content in the Twentieth Century: Collaboration Opportunities for Librarians, Publishers, and Vendors*," released in January 2012. The research report presents highlights of interviews conducted from July to October 2011 with fourteen value chain experts. The White Paper also summarizes results of peer-reviewed publications and proprietary research studies to further characterize the currently fragmented discovery environment. In conclusions, the authors propose cross-sector conversations among publishers, vendors, and librarians to further visibility and, ultimately, usage of the scholarly corpus on the open Web and within library services.

Discoverability Study

In May 2011, **SAGE** commissioned a four-month research study which culminated in a White Paper released in January 2012 at the **American Library Association Midwinter Meeting**. The study was intended by **SAGE** to benefit "the community" of publishers, vendors, and libraries. Research project outcomes included: 1) discussion of best practices emerging in discovery and access of content in libraries; 2) identification of problems that publishers, librarians, and vendors need to resolve; 3) suggestions for some real solutions that can be implemented by librarians and publishers; and 4) further observations for improving discoverability and visibility of scholarly content in the 21st century.

In addition to a review of published literature and commissioned studies, co-authors interviewed fourteen library, pub-lisher, and vendor industry experts. This value chain 'convenience sample' was generated through a concluding interview question: Who else should we talk to, and what else should we read? So the list of authoritative interviewees and information sources "snowballed" organically.

As reported in the **SAGE** White Paper acknowledgements, experts contributed insights from 1) scholarly publishers and their published authors and journal editors; 2) search engine developers, journal database aggregators, and abstracting and indexing (A&I) services; 3) electronic resource management (ERM) and integrated library system (ILS) vendors; and 4) academic librarians and library consortium leaders who advance institutional and multi-institutional discoverability.

During the course of the study, "discoverability" was defined as scholars' capacity to locate relevant content in the scholarly corpus as needed to advance their research and other creative activity."[1] Therefore, structured interview questions explored how publishers, libraries, and vendors could collaboratively advance improved discovery of the peer-reviewed/quality-vetted content that academic publishers produce, libraries invest in, and scholars require at appropriate points in their research workflow.

Analysis of detailed interview notes revealed that, in experts' opinions, improved discoverability depends on a variety of strategic cross-sector strategies:

- placing discovery acceleration tools in familiar Web environments,
- detailed indexing for highly relevant and precise search results, and
- seamless identification and fulfillment user experiences.

Accomplishing these means of improving user discovery results and experiences require heightened cross-sector collaboration. In other words, discoverability and, relatedly, visibility require a holistic "ecosystem" approach among value chain contributors — because each sector is part of a dynamic, "whole," interconnected system of information exchange and knowledge creation.

In this symbiotic ecosystem,

- Librarians manage systems for institutional collection, dissemination, and retrieval of the scholarly corpus;
- Publishers produce and promote authors' work through formats findable on the open Web and in library catalogs;
- Publishers' technology vendors supply e-publication platforms and strategic discoverability solutions; and
- Libraries' technology vendors connect publishers' digital content to online public access catalogs (OPACs) through electronic resource management (ERM) systems and Web-scale discovery services.

Therefore, Web-scale discovery and visibility tools depend on value-added, largely invisible contributions of authors, publishers, libraries, and vendors who compose the scholarly value chain. Traditionally, these content and service providers satisfied complementary roles. Publishers provided gatekeeper services, ensuring peer-reviewed content adjudicated by journal editorial boards. In turn, librarians served as access gatekeepers for the authoritative published resources. However, the Internet has disturbed those comfortable and conventional relationships, thereby necessitating reinvention — and recommitment — of centuries-old partnerships among publishers, scholars, and libraries.

Collaboration Essentials

Peer-reviewed journal literature is a primary source of insight, evidence, authority, and attribution in scholarly communication. Traditionally, libraries ensured discoverability and access through a combination of effective cataloging and classification, open and browsable stacks, A&I tools, reference assistance, research consultation, research education, and other services and programs that improved awareness and usage of authoritative information available in and through libraries. Now libraries must re-discover their role(s) amidst increasingly complex workflows, licensure restrictions, statistics analysis, and return-on-investment (ROI) expectations.

Amidst this considerable uncertainty, companies like **OCLC**, **Serials Solutions**, **ExLibris**, and **EBSCO** are partnering with growing numbers of publishers of primary and secondary content (scholarly corpus and A&I services) to produce simplified, centrally-

indexed content. In turn, libraries are increasingly adopting these Web-scale discovery platforms as the preferred interface for library OPACs because they can facilitate local access through a single index that provides relevancy ranking and other facets for achieving precise search results querying content in all formats, whether licensed, owned, or free. **SAGE** White Paper interviewees recognized that furthering discovery, navigation, and fulfillment experiences requires purposeful conversations and heightened collaborations among all these value chain contributors.

Heightened collaboration among librarians, publishers, and vendors is critically important because, despite a disruptive (and disrupted) information landscape, we share a common purpose: to improve discoverability and visibility, access and delivery, and usage and creation of the scholarly corpus. Analysis of cross-sector expert interview data revealed some initial "conversation starters." To begin, agreement is needed on common standards for metadata standards, information organization, and resource presentation. Therefore, an especially fruitful conversation would initiate cross-platform and cross-publisher investigations to identify best industry practices, further shared standards, and apply researcher behavior findings. In response, online product interfaces and publisher Website designs would conform to (yet-to-be determined) standards and functionalities.

In addition, enhanced community collaboration would better ensure researcher navigation to the "best" version of scholarly content for which they have "rights" through academic affiliation validated by institutional authentication. This collaborative outcome would build on the Open URL (link resolver) navigation technology that shows users their options for obtaining target content, whether from the primary publisher's Website, an aggregated collection of content or other options (such as print holdings), interlibrary loan, or document delivery. Functionality is enabled through a combination of technologies and standards and practices, including **National Information Standards Organization/NISO** and **Knowledge Bases and Related Tools/KBART**.

Other promising initiatives, such as **Open Researcher and Contributor ID/ORCID**, aim to provide researchers and other entities with unique identifiers to associate with their research outputs. Version of record is also being addressed to ensure that researchers can "see"/recognize the various incarnations of a journal article through its life cycle of publication and can locate the authoritative and most recent version of a given work. **National Information Standards Organization (NISO)** has also recommended standard version terms, and **CrossRef** has released a new feature for version validation, **CrossMark**.

Meanwhile, Webmasters are increasingly adopting schema such as HTML to construct/mark up Web pages in ways recognized by major search engines, such as Google and Bing, to improve Search Engine Optimization (SEO).

When these search providers directly access databases structured by standardized schema, they can improve discovery of relevant Web pages. Building upon this capability, within the scholarship realm, **ScholarlyArticle** offers a structured data schema to enable improved discovery of appropriate content through consideration of a variety of unique properties, including publisher, editor, reviewer, genre, reviews, ratings, institution, location, creation date, and modification date, as well as author, title, and source — all value-added signifiers of provenance and authority.

Despite progress-to-date, further cross-industry standards are needed for content file formats, structured metadata quality, and online usage statistics to ensure interoperability among search engines, publisher platforms, and integrated library systems, especially as new models for scholarly communicate emerge. With the aim of furthering exchange and creation of scholarship in the digital age, long-time value chain contributors have highly complementary roles to perform, which will surely extend and inevitably re-invent their traditional roles:

• Librarians understand the research and discovery needs of novice and expert researchers,

• Publishers and editors understand the curation, production, and dissemination of scholarly content, and

• Authors and other scholars understand the disciplinary knowledge aspirations and discourse practices of their fields of study.

In addition, newer value chain contributors — libraries' vendors and publishers' vendors — will certainly seek and find new applications for their expertise and products:

• Libraries' vendors and publishers' vendors understand technological infrastructure of platform, discovery, and organizational tools.

Currently, each of these value chain participants contribute significantly to the cycle of creation, discovery, access, and re-creation that catalyzes and informs production of the scholarly corpus which fuels research and learning. However, none yet sufficiently understand the perspectives — and potential value propositions — of the others.

Future Collaborations

After the **SAGE** White Paper went to press in October 2011, two complementary initiatives were launched, and one discussion paper was released, which promise to further real-world collaborations recommended in the **SAGE** White Paper. In late October, **NISO** announced a new Open Discovery Initiative which aims to develop formal standards and recommended best practices for "next generation" library discovery services using an aggregated index search of a wide range of resources, licensed and free, from multiple providers. Toward this end, a new **NISO** committee will convene open discovery libraries, information content providers, and discovery service providers to advance creation of consistent vocabulary and business practices. One anticipated outcome

is clarification of exactly which resources are available in uniquely licensed and purchased electronic content and which are indexed in full text or by citations only, or both, and whether the metadata derives from aggregated databases or directly through the full text.[2]

In a highly complementary action in February 2012, the **National Federation of Advanced Information Services (NFAIS)** announced a draft Discovery Service Code of Practice for review and comment, in the belief that:

"discovery services have the potential to provide ease of information discovery, access, and use, benefitting not only its member organizations, but also the global community of information seekers. However, the relative newness of these services has generated questions and concerns among information providers and librarians as to how these services meet expectations with regard to issues related to traditional search and retrieval services; e.g., usage reports, ranking algorithms, content coverage, updates, product identification, etc. Accordingly, this document has been developed to assist those who choose to use this new distribution channel through the provision of guidelines that will help avoid the disruption of the delicate balance of interests involved."[3]

To also further conversation, in December 2011, **OCLC** released a discussion document, *Libraries at Webscale*, which presented views of leading thinkers and writers in the fields of information, education, marketing, and technology, who responded to the question: "what next?" They concluded that:

"big collaboration in the information ecosystem will come not only from broader collaboration across libraries, library groups, consortia, and cooperatives, but increasingly through new, innovative alliances and partnerships across the broader knowledge community — across researchers, publishers, commercial vendors, and Web-scale providers such as Google, Amazon, and Facebook."[4]

Noting that an ecosystem thrives through complex relationships and interactions among its members, the document offers several possibilities for building relationships and interactions within a Web-scale information ecosystem:

• Connect users with content regardless of format or where it is stored by creating new models of partnership with all types of content providers,

• Develop new forms of knowledge through dialogue and discourse that are easily distributed, reviewed, and added to the collective collection,

• Build creative "spaces" that encourage collaborations of pure exploration and invention among any ecosystem members, organizations, or groups, and

• Build bridges, links, and tunnels to wells of information that make it easy to find, connect, compare, mix, or mash up all content into any format.[5]

In concluding, the **OCLC** report notes that the network of organisms within an ecosystem contributes to its growth and expansion by facilitating adaption, change, and contribution. A critical balance between cooperation and competition generates energy and motivates the evolution of the ecosystem toward higher function, nourishing the entire community. In a Web-scale world, collaborations must both promote sharing and drive innovation.[6] As demonstrated in the **NISO** and **NFAIS** instance, this will require establishment of shared values and principles that can support cooperation and commerce through partnerships that co-create a vision of the future with content publishers and their platform providers, libraries and their service providers, library consortia, and national and international standards initiatives. "A Web-scale world makes this conversation urgent — and exciting."[7]

Endnotes

1. **Somerville, M. M.**, **Schader, B. J.**, & **Sack, J. R.** Improving the Discoverability of Scholarly Content in the Twenty-First Century: Collaboration Opportunities for Librarians, Publishers, and Vendors. A White Paper commissioned by **SAGE**. Thousand, Oaks, CA: SAGE, 2012. *http://surveys.sagepublications.com/Survey.aspx?surveyid=3431*

2. **Walker, Jenny**. **NISO** launches new Open Discovery Initiative to develop standards and recommended practices for library discovery services based on indexed search. **NISO** Press Release, October 25, 2011. *http://www.niso.org/news/pr/view?item_key=21d5364c586575fd5d4dd408f17c5dc062b1ef5f*

3. **Lawlor, B.** email to list.niso.org, February 1, 2012. For full text, see: *http://info.nfais.org/info/codedraftintroduction.pdf.*

4. *Libraries at Webscale: A Discussion Document.* Dublin, Ohio: **OCLC**, 2011, p. 31. *http://www.oclc.org/ca/en/reports/webscale/default.htm*

5. *Ibid.*, p. 32-33.

6. *Ibid.*, p. 33.

7. *Ibid.*

of his wonderful Celtic music out. I listen to them frequently when I get stressed.

Speaking of stressed, I see that someone on my **Facebook** page noted that Stressed is Desserts spelled backwards! I love palindromes, don't you?

And there was even more music involved in the **Penthouse** Interviews! One of our interviewees was the brilliant **Scott Plutchak** who plays with the **Bearded Pigs**, a band of librarians! We are hoping to get the **Bearded Pigs** to the **Conference** for a small gig in 2013! Unfortunately, **Scott** will not be with

Forward into the Past: Offsite Book Depositories, The Future of Libraries?

by **John D. Riley** (Eastern Regional Sales Manager, BUSCA, Inc.)
<jdriley@comcast.net> *www.buscainc.com*

Open stacks are a fairly recent development that can be traced back to nineteenth-century English and American public libraries when their library collections began to exceed the size of the reading room. Book stacks quickly evolved into a fairly standard form in which the cast iron and steel frameworks supporting the bookshelves also supported the floors, which often were built of translucent blocks to permit the passage of light (but were not transparent, for reasons of modesty).[1]

Previous to open stacks, archival storage was the norm. The current practice of offsite storage can just as easily be thought of as archival storage. Books and other materials are kept in a secure, climate-controlled environment with access limited to individual requests filled by librarians or other library personnel. Archives have been a major component of libraries since their inception, and offsite storage has been used ever since the first libraries were created. Most libraries in Europe still keep books in storage with access only allowed by request after searching a catalogue of their available materials. Perhaps their holdings of incunabula and other rare books or simply the scarcity of many books encouraged the practice.

We have reached a similar situation today with an explosion of information and an inability to house all of it comfortably within reach. In addition, mass digitization has quickly converted tens of millions of books to electronic format resulting in less demand for the printed versions. Between these two irresistible forces libraries now find that returning to the archival model for storage, not just of little used items, but current materials as well, is a viable way to continue growing the collection while re-purposing precious space in the heart of their campuses or in urban settings.

I am one of those people who initially was horrified at the idea of storing most library books offsite or in compact shelving. Roaming the stacks was a pleasure I relished in my college years, but it is not something I do very often nowadays. It has become a rarefied pleasure that has possibly been outweighed by the benefits of "archival" storage: secure and safe storage, climate-controlled atmosphere, and easy location of needed items. Some libraries report that up to fifty percent of books searched in open stacks cannot be located, whether because the item was checked out or, more disconcertingly, because it was stolen or simply misshelved. One archive that I visited recently, the Harvard Depository, has lost only two books in its twenty-six-year history!

In fact, my interest in the subject of archival storage came about from a talk I attended given by **Matthew Sheehy**, Head of Access Services of the **Harvard University Libraries**, where he gave a detailed history and tour of the facility using slides and pictures. The size and scope of this project so amazed me that I later asked **Matthew** for a personal tour. He turned me over to the capable hands of **Patrick O'Brien**, Systems and Special Projects manager of the Depository. **Lee Anne Hooley**, Dark Archive Project and Document Delivery Librarian, was a great resource for details about the journal archiving function of the Depository.

I visited the **Harvard** Depository on a cool March afternoon, and it was a good preparation for entering the temperature and humidity-controlled warehouse that is kept at a constant fifty degrees and thirty-five percent humidity. The Depository is also pressurized from inside to create an outgoing breeze when doors are opened to keep out unwanted intruders such as flying insects. So a cool gust of air greeted us as we entered the towering stack area. Summer is the hardest time for the Depository with the infamous New England humidity forcing the air conditioners and dehumidifiers to run twenty-four hours a day.

On this day the Depository was handling its usual hundreds of requests from the **Harvard Libraries** and many from its Borrow Direct partners: **Brown**, **Columbia**, **Cornell**, **Dartmouth**, **New York Public**, **University of Pennsylvania**, **Princeton**, and **Yale**. The partners have access to each others' catalogues, and patrons can "borrow direct" from participating libraries simply by requesting items from their catalogue screens. The books in the Depository are all in the library's catalogue and can be delivered anywhere on campus within a day. Books ordered by 6:00 p.m. are delivered first thing in the morning. Same-day delivery is also available if ordered early enough in the day. The Depository circulates about 2.5% of its holdings annually, around 215,000 items.

The Depository also acts as a "Dark Archive," not unlike a "Seed Bank" which stores seeds against the possibility of some future calamity. By storing runs of journals for **JSTOR** and others, the Depository provides a physical backup to online journals. In spite of the mass digitization of journals, workers at the Depository deliver many articles electronically after scanning the appropriate journal.

The **Harvard** Depository has found that human rather than robotic retrieval of books works best for them. Employing forklifts fitted with work stations, they can go directly to the box they need and retrieve a single book. Books are grouped by size after bar coding and the