

Scholarly Ecosystem Collaboration Potentialities: A SAGE White Paper Update

Mary M. Somerville and Lettie Y. Conrad

International Society of Knowledge Organization United Kingdom Conference

Abstract

The lifecycle of academic works from idea to investigation -- followed by publication, discovery, access, and usage -- is supported by extensive cross-sector collaboration throughout the scholarly communications ecosystem. However, transformational changes occurring worldwide within the knowledge creation and publication landscape have disturbed traditional divisions of labor and established codes of practice, as well as changed researcher experiences and marketplace products. Therefore, long-standing conventions and relationships among libraries, publishers, vendors, and readers/researchers are now being revisited, renegotiated, and reinvented.

With the aim of furthering collaborative cross-sector conversations, SAGE commissioned a study among scholarly communications 'value chain' experts. Results were reported in January 2012 as a white paper titled *Improving Discoverability of Scholarly Content in the Twentieth Century: Collaboration Opportunities for Librarians, Publishers, and Vendors*. Since then, various commissioned studies, research reports, journal articles, standards initiatives, and white papers have offered further insights into the rapidly evolving scholarly communications landscape.

Additionally, established academic publishers are producing new discovery acceleration tools that anticipate evolving research workflow of novice researcher and expert scholars. Within the larger context of evolving international standards, best practices, business models, and codes of behavior, these new 'value added scholarly environments' predict 'pushing the boundaries' for discoverability and fulfillment of the scholarly corpus, as well as its creation, dissemination, navigation, visibility, and usage, on the open web and within library services.

Introduction

Cross-sector cooperation and collaboration opportunities have received considerable attention in recent years, catalyzed by a common aim among libraries and publishers to significantly advance researchers' capacity to locate relevant content in the scholarly corpus and thereby advance academic progress and other creative activities. SAGE was an early contributor to these discussions on discoverability challenges and collaboration opportunities. In May 2011, the publisher commissioned a research study that produced a white paper, *Improving the Discoverability of Scholarly Content in the Twenty-First Century: Collaboration Opportunities for Librarians, Publishers, and Vendors* (Somerville *et al.* 2011), which intended to benefit the community of publishers, libraries, intermediaries, vendors, readers, and researchers who produce and consume the scholarly corpus.

In this exploratory 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. Using a semi-structured interview methodology, the four-person research team explored discoverability issues with fourteen cross-sector industry experts. The SAGE white paper, issued in January 2012, presents recommendations for cross-sector collaborations among scholarly 'value chain' contributors. These sectors include 1) *primary content publishers*, and their published authors, journal editors, and technology vendors; 2) *secondary content publishers* of abstracting and indexing (A&I) services, and their technology vendors; and 3) *academic libraries*, and their campus communities and technology vendors. Amongst the several collaboration priorities that emerged in expert interviews were agreement that enhanced idea generation requires detailed indexing for highly relevant and precise search results, discovery acceleration tools in familiar Web environments, and seamless discoverability and fulfillment user experiences.

In this paper, highlights from the SAGE white paper will be followed by reflection on initiatives and insights available since publication. A number of new developments in the scholarly literature, commissioned studies, international initiatives, social media, and professional conferences demonstrate cross-sector progress toward improved academic discovery.

SAGE White Paper Highlights and Updates

Disrupting forces in the scholarly communications ecosystem have catalyzed new research workflows and discovery methods (Somerville and Conrad 2013; Taylor & Francis 2013; JISC 2012; Rutner and Schonfeld 2012; Gardner and Inger 2012), and challenged long-held assumptions about scholarly gatekeepers, peer review, knowledge containers, social norms, and document functionalities, amidst the emergence of new business models and customer bases. Changes to this landscape are directly impacting how today's researcher discovers and retrieves academic materials and has created new challenges to the research workflow. In response, libraries, publishers, and their respective vendors are leveraging new technologies and weathering turbulent conditions to satisfy – and anticipate – the expectations and requirements of traditional market shares, as well as new constituency groups.

The following sections contextualize the SAGE white paper and provide updated recommendations through discussion of selected cross-sector agreements and initiatives on international standards and best practices for systems, content, and metadata. These advancements are considered within the small but important body of literature on readers/researchers who are discovering scholarly content through an ever-growing range of pathways (Conrad and Somerville 2013).

Search Quality Essentials

Discoverability requires content to be well indexed and well represented. Since enriched metadata is essential for optimum discoverability, the shared aim among ecosystem contributors is to produce high-quality, accurate data that improves search precision, relevance, and, ultimately, content retrieval. As a consequence, metadata standards have received significant attention recently, in an effort to address the uneven protocols for product data (such as title, author, and ISBN) and semantic data elements (such as keywords, relationships, and subject categorization). Additionally, discovery functions like reference citation linking, while they have significant potential to appreciably further discoverability, depend on compliance with author disambiguation and content identification standards,

which rely on collaboration between publishers, authors, and library vendors.

These various issues for discoverability both in the open web and in library systems have been recently addressed throughout a variety of initiatives. For instance, CrossRef, founded and directed by publishers, directs users to the authoritative version/host of a manuscript.

A complementary initiative advanced by representatives from all areas of the community is ORCID (Open Researcher and Contributor ID), which aims to provide researchers and other entities with unique identifiers to associate with their research outputs. To address the researcher name ambiguity problem, ORCID provides a registry of persistent unique identifiers for researchers and scholars. Widespread adoption and usage by the research community at key workflow and dissemination points – manuscript submissions, datasets deposit, grant applications, patent applications, and faculty records – will support cross-disciplinary, cross-sector linkages across multiple grant awards, clinical trials, scholarly publications, patents, and datasets.

Controls for 'version of record' are also being addressed to ensure that researchers have visibility into 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. The National Information Standards Organization (NISO) has recommended Journal Article Version Terms to standardize the language around article versions. In addition, CrossRef has released a new feature for version validation, called CrossMark, which addresses the problem of multiple versions of scholarly content. Articles exist in a variety of iterations throughout the publication lifecycle (author drafts, pre-print releases, corrected manuscripts, etc.) and are hosted across a variety of online locations (e.g., author websites, institutional repositories, government archives, aggregator collections, primary publisher websites, and more). This makes it difficult to locate the most recent authoritative version of a document, or to ascertain if the document has been updated, enhanced, corrected, withdrawn, or retracted. CrossMark also aims to inform researchers if there have been any updates and direct users to the primary host where authoritative version of the paper.

Meanwhile, webmasters are increasingly adopting schemas such as XML and HTML5 to construct (i.e., mark up) web pages in ways recognized by major search engines and discoverable via many devices. When these search providers directly access databases structured by standardized schema, they can improve discovery of relevant web pages. To standardize the metadata embedded in HTML and PDF versions of an article, a number of cross-sector initiatives and data schemas are gaining participation and advancing discovery. For example, e-journal markup routines have been established by NISO in well-established standards such as the Journal Article Tag Suite (JATS) and emerging standards such as Recommended Practices for the Presentation and Identification of E-Journals (Pie-J). Schemas such as ScholarlyArticle 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.

Routine testing now regularly generates new publisher website design practices that ensure optimum search engine optimization (SEO), measured by assessment tools with increasingly sophisticated success metrics, to accelerate document retrieval and enable

researcher browsing. This includes maximizing SEO for mainstream search engines, such as Google or Bing, as well as social media exposure. Many platform providers that partner with publishers also further discovery through content enrichment, quality assurance, and usability testing, with the goal of hosting online content that is easily found and well presented—whether on a publisher’s website or in a library catalog or whether at home or work.

Web-Scale Discovery in Libraries

With the aim of further improving global search, recent initiatives have convened community conversations to clarify codes of behavior and related discoverability issues. NISO’s Open Discovery Initiative (ODI) was formed in 2011 by members in libraries, content providers, and discovery service vendors. NISO defines discovery services as those library applications that provide a single search box to access a central, pre-indexed database of institutional holdings. The ODI aims to advance technological transparency and data exchange best practices for producers and buyers of pre-indexed library discovery services. The ODI supports this needed cross-sector collaboration as indexed content is derived from journals, e-books, and other electronic information of a scholarly nature from a range of information providers. Proposed recommendations for interactions between content providers, libraries, and discovery service providers will be released in 2013.

In a complementary initiative, the National Federation of Abstracting and Indexing Services (NFAIS) produced a draft code of practice (2012), which aims to establish best practices for the business agreements and cross-sector relationships within production and purchase of discovery services. While ODI aims to cover recommendations for related technologies and metadata, NFAIS guidelines inform interactions between the creators of these services and the content providers whose resources they represent. This wide range of activities includes metadata exchanges, content coverage and display, and product identification.

On the other side of discovery, the OpenURL standard advanced by NISO aims to streamline access and retrieval. In a complementary fashion, a United Kingdom Serials Group (UKSG)/NISO initiative known as KBART (Knowledge Bases and Related Tools) was initiated in 2008 to guide standardizing data and practices for electronic resources management (ERM) knowledge bases that populate library website A-Z lists and link resolvers. These initiatives not only illustrate the wide-ranging interests and activities across the scholarly information community—libraries, publishers, ERM vendors, standards organizations, and platform vendors, among others—but also suggest the complexity of coordinated efforts required to attain current levels of reliability and quality across multiple information flows. In recognition of the need to effect smoother transitions between members of the knowledge base supply chain, Phase II recommendations were released in 2012 to enable provision of higher quality data by content providers. In a complementary fashion, the NISO IOTA (Improving OpenURLs through Analytics) initiative released recommended best practice in April 2013 for OpenURL providers to ensure users are provided with complete and accurate links.

For growing numbers of libraries, earlier investment in ERM systems and associated technologies, such as OpenURL software, paved the way for web-scale discovery services. With the goal of fulfilling local access through a single search-box and unified index, discovery services, such as Serials Solutions Summon, provide relevancy ranking, facets for

drilling deeply into search results, and agnostic access to curated library content in all formats. Content within discovery services can be “purchased premium content, while also integrating additional content from local, institutional, or proprietary repositories, web-available digital resources (e.g., HathiTrust), and open access sources,” (Lustig 2011) including journals, books, aggregated databases, multimedia, and other formats.

Such web-scale discovery services provide researchers with search across vast quantities of locally hosted library silos of information, uncoupled from any specific integrated library system. Furthermore, discovery and content retrieval via institutional accounts can occur by mobile access through apps, like BrowZine, or sites optimized for smartphones, some of which issue ‘vouchers’ for off-campus mobile access. As a consequence, libraries can now replicate the centralized, yet flexible, model of Google’s search interface and speed, content breadth, and quality results, thereby finally addressing the vexing question: ‘if Google can do it, why can’t libraries?’

Researcher Workflow Strategies

A rich understanding of researchers’ ‘jobs to be done’ (Page 2013) can drive user-focused improvements to library, publisher, and open-web channels of scholarly content discovery. Amidst the small but important literature on online researchers’ behaviors and experiences, two recent studies focus specifically on web-scale discovery services. Asher and others determined the importance of relevancy ranking, stating “it seems that one of the most important--and perhaps the single most important--factor in determining which resources students will utilize is the default way in which a particular search system ranks and returns results” (Asher *et al.* 2013). The study also noted the variability among discovery services, including the difficulty in assessing services’ proprietary search algorithms. Lown *et al.* (2013) determined that unified library search involves more than the catalog and articles, although such search queries predominate. Additionally, a small number of the most popular search queries accounts for a disproportionate amount of the overall queries, suggesting the merits of ongoing evaluation of library user search behavior to inform discovery layer customization.

One such user-focused study presented at the 2013 conference of the Association for College and Research Libraries examined representative graduate student workflows while completing a thesis or dissertation literature review. Conclusions based on these advanced social science students’ discovery pathways highlighted opportunities for more nuanced collaborations among scholarly ecosystem contributors, such as developing citation and document management systems *with researchers* “to evolve available products, further workflow integration, and advance researcher adoption” (Conrad and Somerville 2013).

Notably, both established academic publishers and software ‘start ups’ are responding to market demand for such researcher tools that support the ‘jobs to be done’ (JTBD) at various points in the scholarly workflow. Products such as Mendeley, Zotero, ReadCube, Papers, and others aim to provide a number of tools that seamlessly integrate with institutional networks and personal digital libraries alike. Some focus on content discovery and retrieval via simple, low-cost document rental or purchase (DeepDyve; Udini); some mainly aim to serve collection, storage, and sharing of resources (Mendeley; Zotero; ARTstor Shared Shelf); others support discovery and storage as well as custom approaches to organizing e-texts or digital images (Papers; Colwiz); and some try to cover all these plus offer unique approaches to PDF and desktop integration (ReadCube; Utopia).

These products are poised to replace the ‘need to know’ strategies currently employed by publishers, which include alerting services from journal websites, widgets that highlight related or recommended content on related sites, and discipline-specific discussion forums and blogs—all of which serve to enhance visibility, promote discovery and, ultimately, drive usage. Innovative academic products support a JTBD approach to supplementing researcher workflows and are beginning to set a new standard for ‘smart’ tools within primary platforms and aggregated databases as well. This type of approach can be seen where publishers are offering visual browse of comprehensive materials on social science research methods (SAGE Research Methods); delivering authoritative topical discussions and cross-publisher literature review references (Oxford Bibliographies; SAGE Navigator); and contextualizing pre-search of authoritative reference source content (Credo Reference). Further innovation of this kind depends on vigilantly monitoring changing researcher needs and habits (which will inevitably change as discovery and delivery functions evolve) to improve the connections between readers and knowledge.

Boundary Crossing Trends

Despite disruptive technological advances and rapid organizational changes within the scholarly community, the driving missions of academic publishers and libraries remain furthering discovery, access, and usage of scholarly publications and creative work. In that spirit, recent advances in metadata standards, information organization, resource presentation, niche products, and industry practices aim to further researchers’ search outcomes. These advancements illustrate the potential promise for more robust discoverability and fulfillment strategies and products achieved through cross-sector collaboration, fortified by supply chain observance of renegotiated practices, standards, and relationships. **Continued progress will require ‘pushing the boundaries’** to achieve seamless discovery and fulfillment experiences through cross-sector conversations and collaborations conducted *with and for* readers/researchers.

References

Asher, A. D., Duke, L. M., and Wilson, S. 2013. Paths of discovery: comparing the search effectiveness of EBSCO discovery service, Summon, Google Scholar, and conventional library resources. ***College & research libraries***.

<http://crl.acrl.org/content/early/2012/05/07/crl-374.full.pdf+html>

Conrad, L. Y., and Somerville, M. M. 2013. Blazing new paths: charting advanced researcher patterns. *Proceedings of the Association of College & Research Libraries Conference (ACRL 2013)*, Indianapolis, Indiana. Available:
http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2013/papers/ConradSomerville_Blazing.pdf

Gardner, T., and Inger, S. 2012. *How readers discover content in scholarly journals: comparing the changing user behavior between 2005 and 2012 and its impact on publisher web site design and function. Summary edition*. Abingdon, United Kingdom: Renew Training. Available:

<http://www.renewtraining.com/How-Readers-Discover-Content-in-Scholarly-Journals-summary-edition.pdf>

Joint Information Systems Committee (JISC). 2012. *Researchers of tomorrow: the research behavior of Generation Y doctoral students*. London: British Library/JISC. Available: <http://www.jisc.ac.uk/publications/reports/2012/researchers-of-tomorrow>

Lown, C., Sierra, T., and Boyer, J. 2013. How users search the library from a single search box. *College & research libraries*, 74(3,May), 227-241. Available: <http://crl.acrl.org/content/74/3/227.short>

Lustig, J. 2011. *Web-scale discovery services: Information management report*. Burlingame, California: Outsell.

National Federation of Advanced Information Services (NFAIS). 2012. *Code of practice: discovery services*. Available: <http://info.nfaais.org/info/codedraft1312012.pdf>

National Information Standards Organization (NISO). 2013. *Open Discovery Initiative (ODI) survey report: reflections and perspectives on discovery services*. Baltimore, Maryland: NISO. Available: http://www.niso.org/apps/group_public/download.php/9977/NISO%20ODI%20Survey%20Report%20Final.pdf

Page, K. (2013). *Customers/users: don't look at needs: analyze the 'jobs-to-be-done.'* United Kingdom Serials Group (UKSG) conference presentation, Bournemouth, UK. Available: <http://www.slideshare.net/kenchad/jtbd-uksg-april2013kenchad>

Rutner, J., and Schonfeld, R. C. 2012. *Supporting the changing research practices of historians*. New York: ITHAKA.

Available: <http://www.sr.ithaka.org/research-publications/supporting-changing-research-practices-historians>

Somerville, M. M., and Conrad, L. Y. 2013. Discoverability challenges and collaboration opportunities within the scholarly communications ecosystem: a SAGE white paper update. *Collaborative Librarianship*, 5(1), 29-41. Available: <http://collaborativelibrarianship.org/>

Somerville, M. M., Schader, B. J., and Sack, J. R. 2012. *Improving the discoverability of scholarly content in the twenty-first century: collaboration opportunities for librarians, publishers, and vendors*. White Paper commissioned by SAGE. Thousand Oaks, CA: SAGE. Available: <http://surveys.sagepublications.com/Survey.aspx?surveyid=3431>

Taylor & Francis. 2013. *Facilitating access to free online resources: challenges and opportunities for the library community*. A white paper from Taylor & Francis. Oxford: Taylor & Francis. Available: www.tandf.co.uk/libsite/pdf/TF-whitepaper-free-resources.pdf