Don's Conference Notes--An NFAIS Workshop and A Charleston Preconference Seminar

Donald T. Hawkins
dthawkins@verizon.net

Follow this and additional works at: https://docs.lib.purdue.edu/atg

Part of the Library and Information Science Commons

Recommended Citation
DOI: https://doi.org/10.7771/2380-176X.7579

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Discovery for Scholarly Research: Evolving Needs and Services — An NFAIS Workshop

Column Editor’s Note: Because of space limitations, this is an abridged version of my report on this conference. You can read the full article which includes descriptions of additional sessions at https://www.against-the-grain.com/2017/01/v28-6-dons-conference-notes/ — DTH

Researchers are now accessing content through a variety of channels, and discovery services have become more important than ever. NFAIS, the National Federation of Advanced Information Services (http://www.nfais.org), held a one-day in-person and virtual workshop on this subject in Alexandria, VA on June 29, 2016. The workshop began with a review by Simon Inger, Principal, Simon Inger Consulting, of the recent report entitled “How Readers Discover Content in Scholarly Publications” that he co-authored with Tracy Gardner. (See my article on the 2016 NFAIS Annual Meeting in the April 2016 issue of ATG and the accompanying online version for a full summary of the report.)

Some of its major conclusions are:

• Web analytics only show the “last hop”, not the origin of discovery, and they often do not capture either geographical origin or users’ demographics.
• Abstracting and indexing services (A&Is) are still first in importance overall, even though a 4-year trend shows some decline.
• Academic researchers rate library discovery as high as A&Is.
• Students' reading lists are often the first entry point for finding material, but they must go beyond the lists.
• Many factors including brand, ease of use, information literacy training, and availability of resources influence selection of a discovery service.

Discovery Tool Services

Mike Showalter said that data discovery varies; large datasets tend to be easily found, but smaller ones such as those connected to a single article are more difficult.

OCLC has recently produced a compilation of articles on the library in the life of the user; some of its conclusions are:

• Discovery applications are just one tool to use.
• Users’ expectations are driven by what they use in other parts of their lives.
• The technology train keeps rolling; where will it be in 10 years?

In considering discovery, we tend to focus on advanced users, but we must recognize that undergraduates account for a significant amount of the use of discovery services. When those students become graduates, their expectations will be very different than we may think today.

Dan Driscoll, Vice President, Database Partnerships at EBSCO, said that EBSCO’s relevancy ranking involves more than simple keyword matching, and some metadata fields count more heavily than others in scoring. The goal is to determine what an article is about, not just find the keywords. Unstructured and imprecise keyword searching has been replaced with precise concept searching; user concepts are matched with the appropriate equivalent vocabulary terms. “Did you mean” suggestions are a significant advance on spell checking, and EBSCO’s suggestions were significantly upgraded in 2015. EBSCO has also developed a “Research Starters” product based on data from PhDs at Salem Press and Encyclopedia Britannica. Alternative metrics from Plum Analytics (http://plumanalytics.com/) are better than citations and will be added to result lists in EBSCO’s EDS discovery service.

Christine Stohn, Senior Product Manager at ProQuest/ExLibris, noted that discovery is more than searching: it is a gateway that is used in context to guide users to other resources. Users are impatient, mobile, and social; they want simple fast results, will not read long explanations, and do not like cluttered pages. They are accustomed to personalized experiences which are difficult to accommodate in discovery services. Here are some of her conclusions from user and usability studies.

• Discovery is about finding specific topics and going beyond known items and topics.
• Users often consult with peers and start a search with some knowledge of a topic.
• Many users start with Google because they are used to it and find it simple.
• Students’ reading lists are often the first entry point for finding material, but they must go beyond the lists.

A&I Databases in Discovery

Joelle Masciulli, Head of Research Discovery at Thomson Reuters, described the role of Thomson’s Web of Science (WOS) product in discovery. She began by listing some of the top trends affecting research and researchers:

• There is an increased focus on collaboration, especially across disciplines and geographic areas.
• The demand for open science and data will continue to grow.
• Career and reputation management is important everywhere. Researchers need to be sure they are representing themselves well.
• All science is computational, so data must be linked at multiple levels.
• Problem-oriented contextual research with an emphasis on solving practical rather than theoretical problems is growing, which has resulted in a decline in the distinction between science and technology.

The WOS today contains over 62 million high quality records with over 1 billion cited references going as far back as 1898, all of which continued on page 75
are searchable. It is a unique collection of metadata about the research ecosystem that can be accessed as a citation network to reveal connections between scholarly works or to generate analytics. Overall usage has grown significantly in recent years; in 2015, the WOS was the top DOI referrer to CrossRef.

Much of today’s emphasis is still on search, not discovery, so a new “WOS Everywhere” concept provides quick powerful access to the global research ecosystem using the world’s leading citation databases. Data is taken from 12,500 of the highest impact journals in the WOS core database, a new “Emerging Sources Citation Index,” and regional citation indexes from emerging economies. The next step is to further harness the power of the citation network by viewing the connections among researchers in new ways: through ideas, institutions, funders, etc. so that the way researchers engage with the literature and each other, explore connections and new disciplines, and keep current will be transformed. Discovery must come to the user, which will bring a more social experience into the WOS.

Jessica Kowalski, Director of Market Development at Elsevier, said that there has been a decline in usage of A&I products, primarily because new forms of usage are emerging. In the past, discussions of A&I services have tended to focus on a few key players, but today, the research landscape has dramatically expanded, as shown here.

![Image of various search services]

In 2012, the primary decision criteria for selecting an A&I service were the breadth of its database, ease of use, and citation quality; today, the criteria are content coverage, author profile capability, and presence of citation analysis tools. To survive, A&I tools must continue to expand their role in the research workflow. Formerly, they connected the initial search to content; now they must also include information from other sources, such as funding, alternative sources, etc. Disambiguation of resources by author or affiliation, integration with local sources, and analysis of citation data and metrics are all important features for an A&I service to have.

The most frequently used piece of metadata is funding: if you are cited, are you also being funded? Researchers with the highest visibility receive funding. The current emphasis is on more than citations; we are now entering a phase of “publish, be cited and mentioned, or perish.” Article level metrics provide new ways to measure research impact; all records in Scopus have them.

Social Media and Open Access Impact on Discovery

This session featured two products with different pathways to discovery that can complement the traditional services. David Sommer, Co-Founder and Product Director of Kudos (https://www.growkudos.com/), began with a familiar list of today’s information problems, most of which stem from the appearance of over 1 million new publications every year, which in turn results in too much information, many ways to communicate, and many metrics to seek out and analyze. In such an environment, how can researchers understand which communications efforts will help their work to stand out?

Kudos, an award-winning toolkit, provides tools to help researchers, publishers, and institutions increase the impact of their published work, and is used by over 65 publishers and 90,000 researchers. It works by explaining, sharing, and measuring.

- **Explain**: create plain language explanations of publications. Authors create plain language summaries describing what their article is about and why it is important.
- **Share**: create trackable links for sharing. Kudos integrates with Facebook, Twitter, and LinkedIn, so a single post can appear in multiple channels.
- **Measure**: All authors receive a dashboard that lists their articles and shows the metrics and data used to measure the impact of their work.

A recent study of over 4,800 researchers showed that Kudos does work: sharing increased downloads by 23%. The study also revealed that Facebook is used more commonly for sharing work than one might expect, but links shared on LinkedIn are more likely to be clicked.

Dominic Mitchell, Quality Control Manager of the Directory of Open Access Journals (DOAJ, https://doaj.org/), traced the history of DOAJ since its launch in 2003 and its impact on the discovery of OA content. DOAJ now indexes 9,075 journals from 130 countries that have published over 2.18 million articles. In 2015, there were over 1.5 million referrals to the DOAJ; the top referrers are Serials Solutions and EBSCO.

DOAJ was created to provide a comprehensive service listing quality-controlled peer-reviewed OA journals. It is especially valuable to small independently published journals; with its hallmark of quality, DOAJ provides them with a high level of discoverability. Its metadata is free to use and reuse, and it is open to spiders and crawlers, especially Googlebot. It provides a suite of APIs (see https://doaj.org/api/v1/docs) for the development of analysis applications.

Discovery is as important as availability, and greater discoverability will lead to a greater use of OA. Publishers and editors know that DOAJ can be trusted and can be used to show faculty, researchers, and librarians that OA journals can be trustworthy outlets for research. Google refers 35% of its traffic (a huge amount) to DOAJ, which offers much more information about journals than Google does, and it also has a strong presence on large social media platforms.

In 2015, DOAJ was named as one of the 2 most vital sources for the development of open content. It is a charity that is supported entirely by donations from publishers and libraries, so it is vulnerable in terms of funding. Mitchell therefore encouraged publishers and authors to consider supporting DOAJ.

Emerging Discovery Tools

Dan Valen, Product Specialist at Figshare (https://figshare.com/), said that Figshare is a general all-purpose data repository in which one can easily manage research outputs and make them available in a citable, shareable, and discoverable manner. It provides data management for institutions, cloud services for publishers, and simplification of the research workflow.

Figshare supports the FAIR data principles (data must be Findable, Accessible, Interoperable, and Reusable). Metrics are available on all content. Figshare is free for end users and sells its services to publishers and institutions.

Sara Rouhi, Director of Business Development, Altmetric, LLP, said that alternative metrics (altmetrics) unlock opportunities for discovery. Here are some useful definitions:

- **Altmetrics**: any trace of indicator of online behavior: sharing, downloading, saving, commenting, coverage in news media, citations, engagement on scholarly platforms, web analytics, etc.
- **Altmetric.com**: a data science company dedicated to tracking and analyzing the online activity around scholarly research outputs
- **Research output**: any digital object produced in the research life-cycle.
- **Online activity**: any form of engagement with scholarly research.

---

*continued on page 76*
Altmetrics are useful because they accrue in real time and are dynamic, in contrast to the long lag times with journal citations. Here are some important points to consider regarding altmetrics:

- Altmetrics rarely accrue for most research output. Most altmetrics do not track web analytics.
- Altmetrics speak to attention, not quality (sometimes bad articles get a lot of attention!). Reputation management is very important: attention can be positive, negative, or neutral.
- A post-peer review site should be checked carefully because its data can be very qualitative; only an assessment of the actual mention uncovers new audiences, collaborators, and opportunities.
- Blog coverage is particularly interesting.
- Altmetric data are used to listen to and amplify what researchers in the field are saying. They allow a researcher to be collegial.

A User Journey: University Perspective

According to William Mischo, Head, Grainger Engineering Library Information Center, University of Illinois at Urbana-Champaign (UIUC), over the last 30 years, discovery has progressed from “supercatalogs” including A&I services to federated search systems to web-scale discovery systems (WSDS). Now we have hybrid systems (also called “bento systems”) which are a combination of WSDS and federated searching and present results with content grouped by type or material. WSDS extend the OPAC and integrate local content. Delivery is the paramount concern for libraries; users want to get to the full text as quickly and easily as possible, and the gateway function of libraries is becoming more important.

Many studies of user behavior exist, but more evidence-based data is needed. The Illinois Transaction Log Analysis (TLA) and user surveys studied user behavior and found the following:

- Many queries have over 5 search terms.
- Users make very little use of explicit Boolean operators; instead they tend to cut and paste titles, authors, citations, and DOIs into search boxes to formulate their searches. They depend heavily on the article literature.
- Effective and efficient full text delivery is crucial.
- Over half of the searches are for known items.
- Users frequently have a material type in mind when they search.
- The use of search assistance is high.
- Gateway tabs to limit searches to material type are used in about 24% of the searches.

The UIUC library’s gateway portal is powered by its in-house developed Easy Search federated search system (see http://library.illinois.edu) which features contextual and dynamic search assistance and is incorporated into the bento system. Nearly 60% of the searches start from the Easy Search Everything tab; only 4% use the Advanced Search tab. Users like the bento display of results. There is still a need for a display of catalog item availability and direct links to eBooks.

Remaining questions for discovery systems:

- Are bento displays better?
- Should the focus be on known-item searching?
- What is the library’s role in discovery?

Challenges and Opportunities

The final session was a general discussion and summary which produced this list of the major conclusions of the workshop:

- Discovery has solved many problems for publishers by exposing a lot of their content.
- Even if no money changes hands, relationships are still important and worth cultivating.
- Everything on West and Lexis is not discoverable on a discovery system. There is lots of content like that.
- If you are the first one to buy something, you can spend a lot of time creating records for the systems.
- There is much content in which users are interested that is not articles, such as photos, maps, videos, news, etc. Most discovery issues seem to be oriented towards articles.
- Personalization is at a crossroads because of privacy and questions of who the user is.
- How engaging a publisher website is depends heavily on the business model and whether it can get the user to pay something.


This preconference seminar at the 2016 Charleston Conference addressed some growing problems in the publishing industry. Six speakers addressed predatory publishing, piracy, and privacy. Several of the presentations addressed current issues surrounding Sci-Hub, a controversial search engine that provides unauthorized free access to articles outside of publisher paywalls by using access credentials obtained from researchers.

Predatory Publishing

Rick Anderson, Associate Dean for Collections & Scholarly Communication, University of Utah, and current President of the Society for Scholarly Publishing (sponsor and organizer of the seminar), began with an overview of pirates and predators from the viewpoint of the library. He noted that piracy is not new (the first recorded use of the term was in 1603), but it became a major issue in the mid-1990s when information stopped being encoded in physical objects. The Internet has made all copying and dissemination, including piracy, radically easier, and piracy has become very difficult to stop. In addition, a growing dissatisfaction with the scholarly communication economy has given rise to an opposition to paying for content.

Predatory publishing (which is really deceptive publishing) is both old and new, especially in the areas of scholarly monographs and journals. The result has been the appearance of a stream of books with scholarly sounding titles, but with low quality or relevance and of little use to anybody. The Internet has also made it easy for predatory practices to flourish.

Endnotes

6. Detailed reports on many of UIUC’s analyses are available at http://www.library.illinois.edu/committee/dbst/discoveryresearch.html.
publishers to issue journals with little expense or investment of time. These journals are almost always open access (OA), but criticism of predatory publishing should not be construed as a criticism of the OA movement. Predatory publishing is really an outgrowth of the Author Publishing Charges (APC) model which gives publishers an incentive to accept as many articles as possible.

Regina Reynolds, Director, US ISSN Center at the Library of Congress, said that many people are serving on the editorial boards of predatory journals, sometimes unknowingly, and some editors-in-chief never even see the articles before publication! The victims of these practices are libraries, junior authors, and scholars in developing countries who need a place to publish. Reynolds quoted data from a recent study that showed the rapid rise of predatory OA journals in the last five years. From 2010 to 2015, the number of predatory journals grew from about 1,800 to 8,000. Governments are supporting OA which is encouraging, and there are some very high quality OA journals; however, the increasing numbers of researchers has also resulted in a rise in the number of predatory publishers because researchers need a place to publish and many are willing to pay to disseminate their results.

The growth of predatory publishers is also raising the concern of librarians, who are in a good position to call attention to them. Jeffery Beall, a librarian at the University of Colorado, Denver, began publishing a list of possible predatory publishers in 2009. Although Beall’s list has some detractors, it has served a useful function in raising awareness of predatory publishers.

Since one tactic of predatory publishers is to give their journals names closely resembling those of legitimate journals, Reynolds suggested that one way to distinguish predatory journals would be to assign an ISSN to all journals. Each ISSN has metadata behind it to allow a title to be distinguished from similar titles; however, Reynolds cautioned that the ISSN is merely an identifying number that does not indicate quality or legitimacy. The ISSN Center has published guidelines for issuing an ISSN.

Reynolds identified these recent promising developments:

- The ISSN International Centre has developed the ROAD database, a directory of OA scholarly resources, which also provides information on the quality of OA publications and gives a view of global OA scholarly publishing.
- The Directory of Open Access Journals (DOAJ) established best practices and has tightened its criteria for inclusion. As a result, over 3,000 journals were removed.
- The Open Access Scholarly Publishers Association (OASPA) has established principles of transparency and best practices in scholarly publishing.

She also suggested that academia could play the following roles:

- Raise the awareness of predatory practices and low quality journals,
- Assess “publish or perish” requirements,
- Assess the value of OA journals,
- Evaluate journal quality vs. prestige, and
- Scrutinize journal service as editors or editorial board members.

Sci-Hub is a wake-up call for publishers. Its continued existence despite legal actions against it may be a consequence of access difficulties experienced by researchers. Reynolds wondered if Sci-Hub is a result of subscriptions-only access or a symptom of a more general problem with OA. Her conclusion is that we are in transitional times marked by chaotic conditions.

David Crotty, Editorial Director, Journals Policy, Oxford University Press, said that many researchers are deliberately choosing to publish in journals from predatory publishers, a large majority of which are located in Africa or Asia. In one way, they can be viewed as a response to a market demand. Researchers are using them as a path of least resistance to get their work published. Among the factors that authors use in choosing a journal are its reputation, relevance to the discipline, impact factor, and readership. Finding the right audience is a key factor.

Most legitimate publishers have spent a long time investing significant effort to establish a reputation; newer ones are the most likely to be hurt by predatory publishers because researchers tend to be skeptical of a relatively unknown publisher. OA is seen as a growth path by many publishers even though it has been tarnished by predatory publishers. The general public and media cannot easily distinguish between a predatory and legitimate publisher, which leads to an undermining of public confidence in science. Academia must make these issues a priority and set standards with strict requirements for journals. A third party is needed to monitor journal behavior, despite possible resistance from publishers.

Copyright Piracy

According to Craig Griffin, Solutions Engineer, Silverchair Information Systems, indications of piracy include mass downloading, data breaches from traditional hacking, and sharing content on a massive scale. A significant issue in the piracy debate is to determine when sharing becomes piracy: how many people do you need to share with?

Sci-Hub was established in response to high article prices, the need for access through an institution, and the desire for convenience. Every publisher is affected by it. Even the act of searching for an article by a researcher can lead to its inclusion in Sci-Hub’s database: if the article is not already in its database, Sci-Hub uses one of its access passwords to find and download it. Elsevier has spearheaded the legal opposition to Sci-Hub.

Ken Varnum, Senior Program Manager, University of Michigan Library, said that obtaining access to content in the traditional (“right”) way has many advantages. It respects the intellectual property of the author, ensures the long-term validity of the scientific record, provides altmetric credit where it is due, gets value from a service the library pays for and demonstrates its value to management, and provides assurance that the content does not have any viruses.

However, user experiences with online content could be better, which we can easily see if we consider the steps the user must go through:

- Figure out how to start and get on the right network,
- Determine how to log in,
- Decide which link is the correct one that will provide the necessary access,
- If there is a link resolver, find which possible option is the right one for them,
- If links are broken, figure out how to report and solve problems.
- Finally, get the article.

In contrast, consider the “dark side.” It provides instant access to the content through a very smooth user interface, and there is no need to expend staff resources updating entitlement lists or troubleshooting the various elements in the system. There is therefore a strong temptation to access content the easy way, which is efficient but totally wrong.

In either case, we still must be concerned with the user’s privacy, confidentiality, and experience, and be able to help users get the information they need when they cannot find it themselves. The user experience is absolutely critical; if we cannot demonstrate value, we will have a hard time acquiring new technology or new content.

Todd Toler, VP of Digital Project Management at Wiley, discussed universal research access in the 21st century. He said that users need to be able to start their information journey from anywhere and have a frictionless experience on any device with an experience as easy as using Sci-Hub. Publisher business models must be preserved without blocking IP addresses but must also prevent access to systems such as Sci-Hub. And user privacy and personalization must be protected without requiring them to register on publishers’ platforms.

Here are some existing issues:

- Unsecured wi-fi networks are the most vulnerable part of campus computing facilities.

continued on page 78
• Off-campus access is cumbersome.
• Personalization requires registration and authentication on each publisher site, and each publisher has its own unique interface.
• Systems lack scale or an agreed-upon infrastructure between information providers and consumers.

Toler suggested the creation of an institution’s authentication server, to which new users or users with new devices are automatically directed when accessing any publisher’s site. Once they authenticate themselves with that server, they can use their device to access any publisher’s content from anywhere in the world. In this model, the authentication moves from the publisher’s site to the user’s institution, and there is no further registration or maintenance. The servers can use any type of authentication and transmit only the unique user identification and metadata required; no personal data is shared. A “WAYF” (Where Are You From) cloud is an intermediate solution and is still needed. It would provide the connection between the user’s device and publisher sites by installing a cookie on the user’s device and keep track of all information provider systems and the content to which the user has access. The main issues are:

- How to get this model on to a library’s agenda,
- The readiness of institutions to maintain these environments,
- Publisher platforms that support this technology,
- Maintaining the library’s mandate for privacy but still allowing publishers to transparently collect the usage data they need.

Privacy

Todd Carpenter, Executive Director of NISO, said that security and privacy are intertwined. If you want to protect something, you must think about the value of what you are trying to protect. There is no common definition of private information. What you consider private may vary depending on the circumstances, and things you consider private may actually not be. Network analysis might signal something about you; for example, if Amazon’s book cover images appear in a catalog and are clicked on, that might provide an insight into user behavior. We need a better balance between privacy and providing services. The tactics we now use to understand our users are not working very well; opt-in systems might be a possible way to protect users’ privacy.

Closing Summary

After some audience round-table discussions, summaries were presented.

Predators

- Cabells International® will be taking over the production of Beall’s list. They have been publishing a journal directory for 30 years and are well qualified for this task.
- Not everyone has access to an elite university and its information resources. How can we do a better job mentoring junior faculty and help them build a legitimate career?
- What is the role of research societies? Training for peer review?
- What is the role of libraries? Some of them provide literature guides, but it is not clear who reads them. Libraries could perform a useful service by providing ethical training for scientists.
- Anything publishers issue would be helpful for libraries.

Piracy

- Libraries generally do not provide appropriate user education about the use of passwords.
- Communication is necessary when a new security system is implemented.
- Many different parties must work together to combat piracy; how ready are institutions to integrate? (Generally, the larger ones are ready, but the smaller ones are not in a position to make changes.)
- What are good things for the user? Confusion stemming from the need to log in to systems with different credentials should go away, so that search and discovery can be a richer experience. There are benefits to a universal identity system.
- IP address management is a difficulty now for many administrators because access should be based on individuals, not institutions.

Privacy

- Pressures on libraries are forcing them to address privacy issues. They want to deliver good access but also good ROI to administrators,
- Libraries have a long tradition of protecting privacy. Policies are well meaning, but now we have a population that is much more comfortable sharing than in the past. They want to be able to access their own data.
- More experiments are needed; in the future only the data actually needed should be collected.
- Vendors should help provide some level of support to users; more tracking may help them to be better partners with librarians. We need to study this in more detail.

Donald T. Hawkins is an information industry freelance writer based in Pennsylvania. In addition to blogging and writing about conferences for Against the Grain, he blogs the Computers in Libraries and Internet Librarian conferences for Information Today, Inc. (ITI) and maintains the Conference Calendar on the ITI Website (http://www.infotoday.com/calendar.asp). He is the Editor of Personal Archiving: Preserving Our Digital Heritage, (Information Today, 2013) and Co-Editor of Public Knowledge: Access and Benefits (Information Today, 2016). He holds a Ph.D. degree from the University of California, Berkeley and has worked in the online information industry for over 45 years.

Endnotes

3. https://scholarlyoa.com/2016/01/05/bealls-list-of-predatory-publishers-2016/#more-6533
6. https://doaj.org/