

Strategic Access

Abstract: There is a revolution afoot in scholarly communication and it is called, Open Access. Whilst Gregory J Gordon is a strong proponent of Open Access, he believes many people are missing the point as more does not mean better, it only means more. Open Access has had a major impact on scholarly communications by reducing the traditional barriers to research. Unfortunately, this has compounded the issues of discovery. Institutional repositories, disciplinary repositories and multi-disciplinary, multi-institutional repositories and metrics such as downloads, citations, and Eigenfactor™ Score are new tools in the scholarly researcher's kit. Familiarity and comprehension of these tools will help scholars make efficient use of the overabundance of scholarly communications. This strategic access approach will ultimately result in greater precision.

Keywords: open access; repositories

There is a revolution afoot in scholarly communication and it is called, 'Open Access'. While I am a strong proponent of Open Access, I believe many people are missing the point as more does not mean better, it only means more. Instead, we should think about accessing content when and where we need it. We should be accessing it strategically.



Gregory J. Gordon

of a publication deemed of value by a publisher such as price, permission, and availability, a common restriction remains; the attribution of authorship. In other words, published scholarship should be readily available for scholars and researchers to utilize and cite in their scholarship and the author must be credited. Open access advocates argue that the traditional barriers erected by publishers such as price, permissions, and limited access, inhibit the process of scholarly inquiry and ultimately hamper the progress of scholarship and research. With open access citations being 10-20% higher, these arguments are not without merit.³

OPEN ACCESS

Before we can consider the weightier issues of relevance and precision, we need to think about access. Therefore it is crucial to define open access, as there is a decided lack of clarity surrounding this concept. According to Peter Suber, the *de facto* spokesperson of the open access movement, "Open access literature is digital, online, free of charge, and free of most copyright and licensing restrictions."¹ Thus open access scholarship is freely and readily available for reference to existing work as well as use within scholarship in process.

While the concept of copyright has proven to be anything but stable over the centuries, the open access advocates brought the issue to the fore within scholarly communication. The open access movement hinges on three declarations over a 20 month period. The first was the Budapest Initiative in February 2002, then the June 2003 Bethesda Statement on Open Access Publishing, and finally the October 2003 Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. For the sake of simplicity, Suber has distilled the commonalities of these three statements into the "BBB definition of Open Access."² While open access strives to remove most facets

The basic tenets of open access seem quite straightforward, however a variety of types of OA exist. These differences refer to the delivery mechanism of the articles and status of the traditional barriers to access. The chromatic qualifiers – gold and green – indicate if the work is available OA via a journal (gold OA) or by way of a repository (green OA). Thus an article published in *The Journal of Legal Analysis*, an open access journal, is considered *gold OA*. A pre-print article deposited within an institutional repository to be published within a conventional journal available only via subscription is an example of *green OA*. The status of other barriers to access, such as price and permission for reuse, are indicated by the terms *gratis* and *libre*. A *gratis* OA publication is free of price barriers as the publication is openly available, free of charge. This can be achieved in a variety of ways, from the initial funding model to a system wherein publishers charge the author a fee to 'free' the work. However many OA advocates do not consider this a truly *gratis* publication. A publication is considered *libre* if

one or more of the permissions barriers are also relaxed.⁴ Many different permutations of OA are in effect; however the basic goal behind each is to increase access to the scholarly research contained therein.

ACCESSING SCHOLARSHIP

Academe has changed significantly since Tim Berners-Lee and Robert Cailliau invented the first web browser at CERN (European Council for Nuclear Research) in Geneva in 1991. This technological innovation enabled virtual access to the SPIRES-HEP bibliographic database and was the start of web servers accessing a common database for research purposes.⁵ The internet has enabled fast, efficient paper distribution to scholars. Today's online systems provide scholars with the opportunity to search for full-text papers and download these directly wherever they have a connection. Changing the constraint from location⁶ to connectivity is significant because it eliminates physical limitations.

Most publishers and many academic institutions, prompted by this technological revolution, created online repositories for aggregating and providing access to research and related materials. Repositories exist for a variety of reasons and in a variety of forms. Two types of repositories traditionally factor into this discussion: the institutional repository (IR) and the disciplinary repository (DR). While IRs are generally supported by an institution and accessible to the public, DRs require external funding or other means of support, potentially including download and other access restrictions or subscription services.

One of the primary goals of an institutional repository is the centralization of organizational content for archiving purposes. Instead of its papers being scattered throughout the internet on publisher websites, scholarly society conference proceedings, scholars' personal websites (or even worse, their personal computers), IRs provide scholars with a secure place to store and easily share their research. Additionally, the organization can generate greater interest in its scholarship as a whole and a potentially wider audience for its research by branding the repository and presenting a collection of centralized knowledge. Broad exposure and simplified access are key factors for a successful repository. Prominent examples of a successful IR include Oxford University Research Archive (ORA) (<http://ora.ox.ac.uk/>), eScholarship at University of California (<http://escholarship.org/>), and Digital Access to Scholarship at Harvard (DASH) (<http://dash.harvard.edu/>).

Led by librarians and often funded by foundations and similar organizations, these initiatives have, however, not always proven successful.⁷ Despite significant grants and other sources of funding, the volume of content stored in IRs remains relatively small when compared against the volume of research generated by the institution as a whole.⁸ A foundational issue and the cornerstone for a successful repository is ingesting enough content to provide a critical mass in one area or another so that users consider visiting the repository worth their time.

The other major type of repository is the subject-based or Disciplinary Repository (DR), which is built around a specific subject matter. These often arose within a discipline (or developing discipline) lacking a single institution with a critical mass of subject area content. Aggregating content for a specific subject area in a DR can create a well-developed collection of research more quickly, as it is not limited to the output of any single organization. All repositories seek to aggregate and archive content, but DRs have the ability to focus on the needs of their disciplinary community without traditional organizational constraints.⁹ Due to their content, disciplinary repositories tend to offer enhanced services to their community, which in turn stimulates dedication to the collection. Such benefits include the development of discipline-specific tools that allow users to dig deeper into the content, and customized filters providing for more granular access and community-based presentation of the content. By concentrating the knowledge for a given field in one virtual location, users are able to view the discipline as a whole and browse the literature to determine patterns of influence. arXiv at Cornell Library, which focuses on physics (<http://arxiv.org>) and CiteSeerX at Pennsylvania State University, which focuses on computer and information science (<http://citeseerx.ist.psu.edu>) are prominent examples of DRs.

Repositories provide authors with a simplified platform for sharing their research while also increasing exposure for the research. Users of the research generally gain access to a greater depth of content within an institution or discipline through repositories. The core problem has been that IRs and DRs are limited to one institution or discipline. Recently, we have seen these two types converge into the multi-disciplinary, multi-institutional repository (MDIR) such as the Social Science Research Network (SSRN <http://ssrn.com>). Just as a computer is a convergence of multiple technologies – devices that calculate, store information, and embody a degree of automatic control¹⁰ – the synergy of the MDIR is greater than the sum of individual parts.

MDIRs provide the core access and exposure benefits of institutional repositories and disciplinary repositories with an interdisciplinary, inter-institutional twist. These hybrid repositories allow classifications into multiple subject areas across disciplines and institutional categories. For example, a law paper could also be classified as accounting, economics, or finance, giving researchers from these other disciplinary homes, new and innovative, perspectives when exploring their research topic. This model offers exposure to different ways of thinking and concepts, all within the same repository. The cross-pollination of ideas, and discovery across institutions, creates a platform for innovative research. Researchers tackling the same issue from disparate disciplinary perspectives will engage with different facets of the issue and identify solutions, when viewed together, that are far more comprehensive than a traditional law review or journal article.

The MDIR model exposes content across the disciplines and, being a repository, provides simplified access to this information. In many cases it provides access to

multiple stages of a research paper; idea paper, working draft and final version, and multiple versions; conference paper, university series, and published version(s). These different stages and versions allow readers insight into the evolution of the ideas and in many cases experience a more thorough discussion in the electronic version than is possible in the print version.

One opportunity for librarians and other information managers is to use MDIRs to reduce the scholar's time spent doing basic research. Searching for an article not yet published or otherwise inaccessible is frustrating. Trying to look at the problem from a different perspective is extremely difficult and sometimes impossible. An MDIR can provide efficient access to content allowing the scholar to produce innovative research faster.

EVOLVING SCHOLARSHIP

A significant recent change in scholarly communication process is the blurring of the boundaries between working papers, current research in its formative pre-publication state, and accepted papers, final versions that will be published in a peer-reviewed journal. This once clear distinction has eroded considerably. Some working paper series are peer-reviewed and many publishers provide access to pre-publication papers, even after they have been peer-reviewed. Working papers have traditionally been defined as preliminary scientific papers but inclusion in a working (or research) paper series may often be the final stage for a paper. Accepted (or peer-reviewed, refereed) papers are more of a continuum than a cliff. They are made available in various ways along the process from submission to initial review to formal acceptance to online or printed publication. Working and accepted distinctions are fine in certain circumstances but they do not accurately fit the broad spectrum and the overall encompassing term of 'scholarly papers' makes more sense. It also allows for each of the current stages and the developing new ones to fit together.

INFORMATION OVERABUNDANCE

With the blurring of boundaries and broader access to research, a major issue for scholars is managing the overabundance of content. This information overload requires more sophisticated tools to manage this potential liability, recasting it as a major asset. While recommender systems are being developed and their algorithms refined, Article Level Metrics (ALM) provide excellent signals for identifying which papers should be read. Within SSRN, three metrics are available: Downloads, Citations, and Eigenfactor™ Scores.

Downloads of articles within a repository are a timely indicator of interest, especially for developing ideas and junior scholars; these statistics provide information about scholarly impact with an altogether different type of metric. Downloads measure the number of times a paper has been delivered to an interested party. SSRN takes

great care to ensure that download counts are an accurate measure of usage and expends a significant amount of resources to maintain their integrity.

The traditional scholarly metric has been the citation. A citation is a reference from one paper to another paper that helps indicate the influence of the original paper. By taking the time to include the original paper the author is acknowledging its value. CiteReader technology, used by SSRN, scans a full text PDF file and captures the references found in it. Those references are then verified through a combination of technological and human review. The verified references are parsed into smaller metadata fields and then matched against other articles in the SSRN eLibrary. It not only provides interesting data on who is citing whom and how often, but it also creates a research timeline allowing readers to easily go backward and forward in a subject matter. The 'References' and 'Citations' pages are freely available for the reader to follow the flow of the literature within and across multiple disciplines.

The Eigenfactor™ Algorithm provides a methodology for determining the most important or influential authors and papers in a network. This algorithm computes a modified form of the eigenvector centrality of each node in the network under the basis that important nodes are connected to other important nodes. This is the basic concept behind Google's PageRank algorithm. Eigenfactor™ Scores have previously been used to rank scholarly journals and the scores are freely available at <http://www.eigenfactor.org>. SSRN uses article level citation data to extend the Eigenfactor™ Algorithm to the author level and will apply it to the paper level in the near future. CiteReader calculates the number of times each paper in the SSRN eLibrary database has been cited by other papers in the eLibrary. This data is then used to construct an author citation network, where each author is a node.

On a technical level, the Eigenfactor™ Score is the outcome of two conceptually different but mathematically equivalent stochastic processes. The first process is a simple model of research in which a hypothetical reader follows chains of citations as she moves from node to node, ad infinitum. An author's Eigenfactor™ Score is the percentage of time that she spends with this author's work in her random walk through the literature. The second process is an iterated voting procedure. Each author divides one vote equally among those authors she cites. In subsequent rounds, each author divides her current vote total, as received in the previous round, equally among those authors whom she cites. This process is iterated indefinitely until a steady state is reached where the number of votes doesn't change. An author's Eigenfactor™ Score is the percentage of the total votes and represents their influence in the network.

Simply put, article level metrics are recommendations for what to read in the increasing world of information overabundance. They purport to serve a similar function to the traditional peer-review filters of scholarly journals – but without the inherent costs and limitations.

MOVING SCHOLARLY COMMUNICATIONS FORWARD

The future landscape of scholarly communication is uncharted. Based on its continuing value over the last few centuries of scholarly communication, we can likely predict an upward trajectory based on the theme of enhanced access to new scholarship. However, increased access brings its own challenges, and herein the focus shifts to the strategic access and management of this content.

First, there are a number of channels that need to be considered. The recent Finch Group Report¹¹ outlines the current environment and recommends UK policy move toward supporting gold OA. It identifies the three primary channels for scholarly publishing, subscription-based journals, open access journals, and repositories:

- Subscription-based journals are published by a wide range of commercial and not-for-profit organisations, including many societies. The publishers sell subscriptions to their journals and a primary concern is that no institution can afford subscriptions to all of the journals;
- Some open access journals charge a fee to authors to compensate for the lost subscription income. Access to the content is free of charge. The number of open access journals has grown significantly in recent years for a variety of reasons;
- Repositories do not act as publishers themselves. They provide access to different versions of papers before they are submitted to a journal or after they have been published.

Second, there is a change in the scholarly delivery mechanism. The core of legal scholarship has been, and continues to be, the law review article. It is a self-contained object; easily transported and comprehensive with its extensive footnotes and references. The footnotes expand the ideas and arguments in the paper. The references provide links connecting the paper to the research upon which it is based. Law review article length has grown significantly in the last forty years¹² and in line with the number of footnotes, often including hundreds of them, until 2005 when law reviews started limiting their length. While some blame the change on digital tools that allowed fast, zero-cost submission to large volumes of law reviews,¹³ I would argue there was a broader, yet subtler change afoot; the evolution of online access and linked content traditionally included in the footnotes. If the legal scholar has been exposed and provided access to this

formerly cloistered information, then the value of the 100+ page, 500+ footnotes massive law review article is severely mitigated.

Third, the process is evolving. The traditional scholarly process is to read papers published in law reviews and peer-reviewed journals in search of an idea either along a current research theme or one that is entirely new. Once the kernel of an idea is found, it must be refined into a manageable and researchable topic. While obvious, this step defines the success of a research project. If a topic is too broad, then the volume of information is likely to be overwhelming; yet if the topic is conceived too narrowly, then the project can quickly grind to a halt. The next step is to thoroughly research the topic using the tools available to you. Using this research, a preliminary draft or working paper can then be written. After several revisions and input from other scholars, the 'idea' is finally ready to be submitted to a law review, conference, or journal. Scholars today are familiar with the internet, but they often lack the research skills needed to take advantage of this evolving world.

The 2009 Joint Information Systems Committee (JISC) paper 'open access – What are the economic benefits? A comparison of the United Kingdom, Netherlands and Denmark,'¹⁴ noted that an author, on average, spent 90 to 100 hours to write a journal article. The authors estimated that a simplified, more open access approach could save scholars 5–10% of the research time.

CONCLUSION

We are in the midst of a period of incredible change, some would argue revolutionary change, and we know open access matters. Open access legal scholarship receives over 50% more citations than non-open access papers¹⁵ but, more is not better and it is not enough.

Scientists, publishers, societies and countless others are looking at open access with excitement and trepidation. It is critical for each of us to think strategically. Our delivery channels and vehicles need to be evaluated and evolved. The scholarly research process that worked for the last few hundred years will need continual updating during this time of change. Article level metrics, when properly utilized, will help us mitigate the inevitable overload of information.

While no one knows how scholarly communications will look in ten years, we do know that strategic access to the research will help scientists create innovative research faster. And, wouldn't all of us benefit from faster research?

Footnotes (with reference to the bibliography below)

¹ Peter Suber, *Open Access* (Cambridge, MA: The MIT Press, 2012), 4.

² Suber (2012) 7.

³ Kim (2012) 1.

⁴ Suber (2012), 6.

⁵ Ceruzzi, Paul, *Computing: A Concise History* (Cambridge, MA: 2012), 133.

- ⁶ Geography, especially in underdeveloped countries, continues to be an access barrier but in this context, location refers to the where the scholar is accessing the content (i.e., in the library, in her office, etc.).
- ⁷ Foster 2005
- ⁸ Nelson 2009
- ⁹ Davis 2007
- ¹⁰ Paul Ceruzzi, *Computing: A Concise History* (Cambridge, MA: The MIT Press, 2012), xi
- ¹¹ Finch Group report
- ¹² Church 1989
- ¹³ Goldman 2005
- ¹⁴ Houghton 2009
- ¹⁵ Donovan 2011

Bibliography

- Boulton, Clint. *Facebook Passed Google, Yahoo, Microsoft in User Engagement*. 29, 2011. <http://www.eweek.com/c/a/Web-Services-Web-20-and-SOA/Facebook-Passed-Google-Yahoo-Microsoft-in-User-Engagement-384309/> (accessed 12 19, 2011).
- Budapest Open Access Initiative. <http://soros.org/openaccess> (accessed July 11, 2012).
- Church, W. Lawrence. "A Plea for Readable Law Review Articles." *Wisconsin Law Review*, 1989.
- CIBER, University College London. *Social Media and Research Workflow*. London: Emerald Group Publishing Ltd., 2010.
- Davis, Phillip M., Matthew J. L. Connolly. "Institutional Repositories; Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace." *D-lib Magazine* 13, no. 3/4 (March/April 2007).
- Donovan, James M. and Watson, Carol A. *Citation Advantage of Open Access Legal Scholarship*. March 4, 2011. <http://ssrn.com/abstract=1777090> (accessed July 11, 2012).
- facebook Statistics. <http://www.facebook.com/press/info.php?statistics> (accessed 12 19, 2011).
- Finch, Janet. "Research Information Network." June 17, 2012. <http://researchinfonet.org/publish/finch> (accessed July 8, 2012).
- Foster, Nancy Fried, Susan Gibbons. "Understanding Faculty to Improve Content Recruitment for Institutional Repositories." *D-Lib Magazine* 11, no. 1 (January 2005).
- Goldman, Eric. "Goldman's Observations." *Law Review Article Length Policies*. 3 22, 2005. http://blog.ericgoldman.org/personal/archives/2005/03/law_review_arti.html (accessed 12 19, 2011).
- Gordon, Gregory J. "The Digital Lawyer's Evolving Education in Scholarly Research." In *Educating the Digital Lawyer*, by Oliver Goodenough and Marc Lauritsen. New Providence, NJ: Matthew Bendert, 2012.
- Houghton, John. *Open Access: What are the Economic Benefits? A Comparison of the United Kingdom, Netherlands and Denmark*. 6 22, 2009. <http://ssrn.com/abstract=1492578> (accessed 12 19, 2011).
- Kim, Heekyung. "MIS Speaker Series." *Eller College of Management, University of Arizona*. January 24, 2012. http://mis.eller.arizona.edu/events/speakers_series/2012/mis_speaker_series_Heekyung_Kim.asp (accessed June 18, 2012).
- Leiner, Barry M., Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, Stephen Wolff. "Brief History of the Internet." *Internet Society*. 12 19, 2011. <http://www.internetsociety.org> (accessed 12 19, 2011).
- Nelson, Bryn. "Data sharing: Empty archives." *Nature* (nature publishing group), September 2009: 160–163.

Biography

Gregory J. Gordon is President and CEO of Social Science Research Network (SSRN), a leading multi-disciplinary online repository of working and accepted paper research in the social sciences and humanities. Currently the number one repository in the world, SSRN provides a variety of electronic distribution and related services to help scholars create innovative research.

SSRN is focused on the high quality, rapid, electronic delivery of scholarly research at the lowest possible cost. More importantly, SSRN is working with scholars to find innovative ways to reduce the researchers' time finding relevant material, provide easy access interdisciplinary content, and accelerate the cycle of research. Its eLibrary database has close to 400,000 papers from over 190,000 authors and users have downloaded over 54,000,000 full text papers since inception.

Prior to helping Michael C. Jensen found SSRN in 1994, he worked at KPMG and two entrepreneurial companies in technology and health care. Gregg speaks around the world and writes regularly about scholarly research and the changes needed to create innovative research faster.