Web Citations: Tracking Citation Impact

Jane Garner, Lynne Horwood and Shirley Sullivan

Introduction

This paper discusses the Web Citation Index (WCI) and other options for citation tracking, such as Google Scholar and Scopus. These are tools that can assist researchers to measure the quality of their papers to support the Australian Government's Research Quality Framework (RQF) requirements. One measure of establishing quality of the research is to indicate metrics, such as the journal impact factor. Another means of establishing peer evaluation is for researchers to be aware of those papers that are citing their own work. This paper provides a commentary about different tools that are available to the research community. Librarians need to know what these tools are to be able to present a comprehensive suite of products to researchers.

Measuring research output

Measurement of research excellence and quality is an issue which has increasingly interested governments, universities and funding bodies, as measures of accountability and quality are sought (Steele 2006b). The RQF, to be introduced in 2008, will be assessing not only the quality of research but also its impact, with 'impact' being a contentious aspect, as it is more difficult to quantify than measures such as the number of publications and research grants. The Australian RQF "differs from existing international research assessment methods by considering research impact in addition to the more conventional quality measures normally used in the academic community." (Duryea 2007) In 2005, the Australian government RQF Research Impact Working Group provided the following statement describing how publicly funded research would be assessed for its impact.

> Research will be assessed as to the extent that it has influenced social, cultural, economic and/or environmental outcomes for industry, government and/or other identified communities, regionally, nationally and/or internationally. (Australia. Department of Education, Science and Training 2005)

It may be helpful to explain further the various uses of the word 'impact' in this paper. "'Research impact' is defined within the RQF as the beneficial application of research to achieve social, economic, environmental and/or cultural outcomes. This is not to be confused with impact in the academic domain, which is seen more as an indicator of the intrinsic quality of the research on scholarly or academic measures. The RQF, in its proposed form, will allow both of these considerations to coexist."(Duryea 2007)

When used in the phrase 'journal impact factor', it refers to a means of evaluation of a journal where the impact factor is defined as "the number of citations to a journal's articles published in the previous two years divided by the number of articles published by that journal during those two years." (Walter 2003)

The Australian Government in the RQF uses the word ‘impact’ in the two ways described above. This paper is primarily concerned with measuring quality in the RQF. When the authors use the term ‘impact’, it refers to the definition in the RQF. When the authors use the term ‘journal impact factor’ it refers to the bibliometric measure used as an evaluative tool for journals. When the authors use the term ‘citation impact’, it refers to the number of times a citation appears in later publications and the implied significance of the cited article.

The Australian government established a Quality Metrics Group to provide advice on the use of quantitative measures to assist in the assessment process for the RQF. This Group focused on citation data as well as grant income data among a range of metrics. (Australia. Development Advisory Group for the RQF 2006)

Rowlands and Olivieri have stated that citation impact should be viewed more broadly and should include many measurable outcomes of research, such as ”whether it is read, cited, used in other research, practically applied, patented, or whether it helps its authors to garner prizes and awards”.(Rowlands 2006)

One method of providing metrics for proving the influence of research within a peer group is by citation tracking. To gain a comprehensive view of the influence of research, citation trackers must be able to index the published journal literature and research housed in institutional and subject based repositories. Web of Science (WoS) has provided a means of tracking citations in the journal literature. With the growth of open access repositories for publishing research findings, there is a need to be able to track the citations of works housed in these archives. Pringle, as quoted by Chillingworth, has said that “the OA and institutional repository community needed a serious index and search tool to make their content more ‘discoverable’." (Chillingworth 2005)

The number of citations an author receives determines, in part, the citation impact of his or her work and the author’s reputation in the discipline. The journal impact factor is defined above as “the number of citations to a journal's articles published in the previous two years divided by the number of articles published by that journal during those two years.” (Walter 2003) Such a measure of impact, although contentious, can be important in attracting funding and obtaining tenure at research based institutions. The journal impact factor’s true objective consists in measuring the international influence or impact of a journal within the research community. As highlighted by Garfield (1999), there is a bias when comparing impact factor of journals relating to different categories in the *Journal Citation Report* (*JCR*) partly because different disciplines have widely differing citation practices. (Figueredo 2006) The journal impact factor was created with the intent of comparing journals*,* not authors or individual articles. If the journal impact factor is used as a tool to evaluate the quality of the individual investigator’s scientific productivity, it needs to form part of a combination of indicators.

Open access repositories

Institutional repositories form a key component of the open access movement to bring scholarly research onto the open web (Quint 2006). The Australian government has recognised their value by supporting two major repository initiatives: One is ARROW (Australian Research Repositories Online to the World). Its objectives are to identify and test software to support best-practice institutional digital repositories and to develop and test a national resource discovery service. The other is APSR (Australian Partnership for Sustainable Repositories), which, as the name indicates, stresses sustainability issues.

The government recently demonstrated further support for open access repositories. The Australian Research Council is encouraging the researchers it funds to make the results of such funded research available in an open access repository (ARC 2006). The National Health and Medical Research Council followed suit with its announcement on dissemination of scientific results.

> To maximise the benefits from research, findings need to be disseminated as broadly as possible to allow access by other researchers and the wider community. The NHMRC encourages researchers to consider the benefits of depositing their data and any publications arising from a research project in an appropriate subject and/or institutional repository wherever such a repository is available to the researcher(s). (NHMRC 2006)

Open access publication models for scholarly communication enhance the dissemination of research findings to all potential users, increasing the economic and social returns to public investment in research and development (R&D). (Houghton 2006) Not only does society as a whole benefit from open access through more effective access to information and an expanded and accelerated research cycle, but the visibility, usage and influence of the work of individual researchers increases (Richardson 2006). Studies consistently show the positive link between citation impact and open access. (Houghton 2006)

Bergstom and Lavaty refer to a number of articles which demonstrate the increased citation rate of open access articles compared to those hidden behind subscription barriers. He notes that articles on open access receive at least 50% more citations than those that are subscription based. Open access articles are cited significantly earlier and more frequently than those that are not open access. An interesting observation is that “self-archiving has become the norm among authors who publish in top economics journals.” (Bergstrom 2007) Self archiving occurs frequently in research intensive institutions and with authors whose papers are accepted in journals commonly accepted as top ranking in the economics field (Bergstrom 2007).

As Pringle describes them, “repositories represent ways of organising research and are taking shape in a variety of experimental forms. They vary in the types of content, the purposes of their creators, and their relationship to researchers.” (Pringle 2005) Different institutional repositories have different

goals. An educational or research institution may use them to collect a corpus of their staff's endeavours and as a showcase of their knowledge output. A government repository may exist as a tool for the dissemination of information to the general public. The content of these repositories includes formats such as technical reports, theses, journal articles, preprints and proceedings.

The reputation of institutional repositories has been increasing since their inception. The fact that they are now deemed a necessary and valuable means of publishing research output is evidenced by their prevalence and by the acknowledgement of many mainstream publishers that authors expect to include their work on their institutional repository as well as having it published in a journal. The Open Archives Initiative and the development of open source archiving software have been the driving forces behind the rise of institutional repositories. As Martello states, institutional repositories are

> a compelling response to two strategic issues facing academic institutions, namely how to reform and complement the current scholarly journal publishing system and how to present to the public an indicator of the quality and relevance of an institution's research efforts. (Martello 2006)

Web Citation Index (WCI)

Thomson Scientific reacted to the growth of institutional repositories by investing in technical development work during 2005 that resulted in the launch of its Web Citation Index. (Swan 2006)

Currently in beta stage, five major institutions were involved in developing a pilot version. The institutions are the Australian National University, Cornell University, the Max Planck Society, Monash University and the University of Rochester.

WCI is designed to be a comprehensive index of scholarly literature within institutional and discipline based repositories, assigning similar tags and indexed fields to this content as it does to the published version of articles from journals. Using WCI, users can search the full text of the repositories, and cited reference search/navigation is enabled where cited references can be harvested from the repositories. The published version of an article is indexed in Web of Science, the preprint or postprint version accessible in an institutional repository is indexed in WCI. (Scholtze 2006) This database sits within the Web of Knowledge platform. WCI bridges the world of the institutional repository with the published literature in the Web of Science and other Web of Knowledge applications. If the IR article is published in the literature and in WoK, a bidirectional link will appear in the full records of each application allowing the user to move between the IR and published view of the record. The WCI has many of the same capabilities as the Web of Science. Subscribers to both tools will be able to perform general searches, search and view cited references, see related records, and view citing records from repositories and journal literature. They can also cross link between these two resources. The WCI is cross searchable and users can use the

Analyze Tool to reveal patterns and trends in their WCI search results. It has weekly updates and will also provide alerts and search history access.

Thomson Scientific has established an editorial team to assess material to be indexed for conformity to a set of selection criteria available at http://scientific.thomson.com/free/essays/selectionofmaterial/wci-selection/. The criteria include the ownership of the archive in question and how well the archive is maintained. Full text availability is sought and document formats are considered. “Repositories must be authoritative, robust, stable, and appropriately formatted.” (Thomson Scientific 2005) When selecting repositories for inclusion in the WCI, Thomson Scientific also assesses the overall design, maintenance and usability of the repositories. The editorial team further considers the frequency of updates and the existence of a review policy or procedure. All records for documents found in the WCI contain well developed abstracts and detailed bibliographic records that include title, author, year of publication and name of home repository. There are currently over 70 repositories in the WCI with an additional 500already selected for inclusion. OpenDOAR and other institutional repositories listings have been taken as a starting point. (Scholtze 2006)

Scopus

Scopus is a database of research literature, including over 250 million scientific web sources. Released in late 2004, Scopus was created from records extracted from Elsevier abstracting and indexing databases such as Geobase, Biobase, Embase enhanced with citation information (Fingerman 2006). Life and health sciences and biological, agricultural and environmental sciences are most heavily represented in the database, followed by chemistry, physics, maths and engineering, with representation from social sciences, psychology and economics. (Fingerman 2006) Though Scopus has records dating before 1996, the citations for a paper will only appear if dated from 1996 on. This may result in many fewer numbers of citations for the same paper found in Web of Science. (Fingerman 2006)

Scopus covers over 15,000 titles from more than 4,000 publishers and includes open access and online-only journals. In 2004 Elsevier established an international Content Selection and Advisory Board comprising independent experts from all fields of science, to evaluate and select information sources for inclusion in Scopus. (Kisjes 2006) International coverage is extensive. Scopus claims that over 50% of its journal titles are from Europe, the Middle East, and Africa. A recent as yet unpublished analysis by Scopus has indicated that 15% of the top 160 journals in which Australian researchers have published between 2003 and 2005 are non-Thomson Scientific journals. (Steele 2006b)

Scopus content is not limited to peer-reviewed journal articles. Scopus coverage extends to 13 million patents and book series. Also important is its coverage of repositories such as arXiv and Cogprints, since many scientists in certain disciplines such as astrophysics, astronomy and physics use those sources extensively. (Fingerman 2006)

In a January 2006 press release, Scopus announced its Citation Tracker, which enables subscribing researchers to evaluate research by using citation data. Scopus describes the feature as offering

> at-a-glance intelligence about the influence of a set of articles, an author or group of authors over time, so users can quickly spot trends. Users can control the specific articles and date ranges they want to evaluate and easily navigate through the cited and citing literature using a visual table of citations broken down by article and chronology.(Scopus 2006a)

In June 2006, the introduction of the Scopus Author Identifier enables users to distinguish automatically between authors with the same name and match variations of author names. (Kisjes 2006)

In August 2006, Scopus announced Selected Sources. (Scopus 2006b) Selected Sources opens the Scopus search engine to subject-based and institutional repositories, comprising courseware, standards, theses, lecture notes, presentations, manuscripts and prepress papers. It provides subscribing institutions with a fully customizable feature that allows users to choose from a list of institutional resources and special subject collections to be made individually searchable. (Scopus 2006b) The results are presented in a simple, tabbed format so the user has all the results in one place. (Kisjes 2006)

While Scopus is searching only about 20 repositories at this stage, owners of repositories can approach Scopus to have their repositories indexed free of charge. The coverage of Scopus is restricted to life, social, physical and health sciences (Scopus Content Coverage 2006), so not every repository will fit within Scopus' coverage guidelines. Examples of repositories currently indexed by Scopus include the institutional repositories of Massachusetts Institute of Technology, Australian National University, Cranfield University, Cornell University, University of Washington, University of Calgary and the Texas A&M University.(Elsevier 2006)

In September 2006, Scopus announced further new features. (Scopus 2006c). PatentCites allows users to track how primary research is practically applied in patents. WebCites enables Scopus users to track the influence of peer reviewed research on web literature. In WebCites, users will be able to view citations to articles in Scopus from a growing number of carefully selected scientific web sources such as institutional repositories and thesis and dissertation databases. (Scopus 2006c)

Google Scholar

Google released a beta version of Google Scholar in November 2004 with the aim of enabling a broad search across the many disciplines and sources of academic and scholarly literature. *Google Scholar* covers print and electronic journals, conference proceedings, books, theses, dissertations, preprints,

abstracts, and technical reports available from major academic publishers, distributors, aggregators, professional societies, government agencies, and institutional and subject based repositories.

Google Scholar's coverage does have serious limitations. In contrast to TS and Elsevier, Google does not offer a publisher list, title list, document type identification, or any information about the time-span or the refereed status of records in *Google Scholar*. (Swan 2006). Elsevier is the most notable absence from Google Scholar's coverage. According to Cockerill, Elsevier

> controls more than 20% of the scientific research publishing market, (and) does not allow any of its articles to be indexed by Google Scholar. As a result, Google's Scholar's coverage is significantly incomplete and skewed. (Cockerill 2006)

Although *Google Scholar* does not cover material from all major publishers (e.g., American Chemical Society and Elsevier), it contains citations to articles from the American Chemical Society and Elsevier when documents from other sources cite these articles. (Meho 2006)

Like WCI and Scopus, it enables researchers to find articles that cite another article and to locate the most highly cited article on any topic. Unlike WCI and Scopus, however, it is freely available to all web users. While Google Scholar is a good tool for finding scholarly articles on practically any subject, its poor structuring and identification of data elements and basic problems with the search engine make it a disappointing tool for its citation analysis function. (Jacso 2006) Other disadvantages include duplicate citations (i.e., counting a citation published in two different forms, such as preprint and journal article, as two citations) as well as the lack of any information about document type, document length, and the refereed status of the retrieved citations. (Meho 2006)

WCI and Scopus could increase the use of repositories significantly by increasing the profile of the IR within the institution. The reputation of an institutional repository can also be enhanced by its coverage by tools such as WCI.

The ability of Thomson Scientific to integrate findings across both the journal literature (with WOS) and the institutional repositories literature (with WCI) is its greatest strength. WCI also provides enhanced reporting methods and graphical demonstrations of citation impact. Features that are available in the Web of Science, such as the ability to organise literature around cited references and the ability to track the relationships among multiple papers, are available in the WCI.

Both Scopus and Web of Science allow researchers to look forward in time as well as back, for example, to see where their own works are continuing to be referenced tracking their own influence, and therefore the influence of their organization. (Fingerman 2006).

Citation Tracking

Although there are moves on many fronts to improve citation tracking, there are limitations to the practice. Many papers have an impact on the research of others without the new work generated being published in traditional sources;

- the IF does not take into account self-citations, which amount to a third of all citations;
- the automatic assumption of a positive link between citations and quality may be questioned as papers may be cited as examples of flawed or suspect research (Walter 2003)

Colin Steele has alluded to another of the problems with citation counts as a measure of quality in that authors are able to manipulate citation counts by self citing or cross referencing the work of colleagues (Steele 2006a). He refers also to the concerns of a leading bibliometrician, Adrian Van Raan of Leiden University, regarding the 'fatal attraction' of Thomson Scientific's metrics as "proxies of research excellence for university league tables and research assessment exercises'. (Steele 2006a)

Another concern has been expressed by the Council for the Humanities, Arts and Social Sciences that:

> conventional Thomson ISI bibliometrics not be considered appropriate measures of quality in the disciplines of Political Science and History at Australian universities because they capture only a small proportion of published articles (about 20%). (CHASS 2006)

Conclusion

Citation tracking has become an essential technique in determining the research quality or impact of an author, or his/her research, on the research community and also the broader community. Authors and their institutions need to be able to determine their influence on the research outputs of others. To gain a comprehensive view of this, there need to be reliable methods of judging the citation impact of works regardless of where they are published and regardless of where their citing papers are published. The rise in importance of institutional and subject-based repositories are increasing in prestige as traditional journal publishers allow authors to place copies of accepted articles into repositories. Institutional and subject-based repositories store vast numbers of documents, including articles, reports, preprints, postprints and conference proceedings. Each of these documents is citing the research of others and being cited by new research. It is imperative that the tools that are being developed to track the citations to and by these documents are comprehensive, deep ranging and broadly focussed.

Although there are many problems with using citation counts as a measure of research impact, inevitably researchers and their institutions are going to need a method of determining such an impact for compliance with the Research Quality Framework and in their applications for research funding and tenure.

Researchers will be relying on citation tracking tools to help them determine their own impact and also the impact of their own work.

A comparison of Web Citation Index, Scopus and Google Scholar reveals that no particular tool is ideal, especially in terms of depth and scope, but each tool mentioned provides a valuable resource. No one of the three tools investigated in this paper will currently provide the research community with a truly comprehensive view of citation impact. When fully operational, WCI should provide a complementary and powerful tool as a part of the Web of Knowledge suite. One of its greatest strengths is its association and linking ability with the other tools within the Web of Knowledge platform. Institutions that already subscribe to other tools within the Web of Knowledge could extend their subscription to include WCI.
Those institutions unwilling to commit to the substantial funding required to subscribe to the WCI or Scopus will be well served by Google Scholar but must be aware of its limitations in scope and depth.

References

Australia. Development Advisory Group for the RQF 2006, *Research Quality Framework: Assessing the Quality and Impact of Research in Australia: The Recommended RQF,* Australia. Dept. of Education, Science and Training, October 2006, viewed 1 December 2006, <http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/research_quality_framework/rqf_development_2006.htm#The_Recommended_RQF>

Australian Research Council 2006, *Discovery Projects: Funding Rules for funding commencing in 2008,* viewed 8 December 2006*,* <http://www.arc.gov.au/pdf/DP08_FundingRules.pdf>.

Bergstrom, T and Lavaty, R 2007, *How often do economists self-archive?* Viewed 20 February 2007, <http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1203&context=ucsbecon>

Council for the Humanities, Arts & Social Sciences (CHASS) Bibliometrics Project, Political Science and History Panels 2006. *Report on Recommendations and Major Issues*, viewed 8 December 2006, <http://www.chass.org.au/papers/bibliometrics/CHASS_Report.pdf>.

http://www.chass.org.au/papers/bibliometrics/CHASS_Report.pdf

Chillingworth, M 2005, 'Thomson herds open access into single index: Web Citation Index creates a global listing of institutional repositories and open access articles', *Information World Review* 23 Nov 2005, viewed 28 September 2006, <http://www.iwr.co.uk/information-world-review/news/2146510/thomson-corals-open-access>.

Cockerill, M 2006, 'Business models in open access publishing',*Open Access: Key Strategic, Technical and Economic Aspects"* Neil Jacobs (ed.), viewed 28 September 2006, <http://demo.openrepository.com/demo/bitstream/2384/2367/4/businessmodelsinoa.pdf>.

Duryea, M, Hochman, M and Parfitt, A 2207 'Measuring the impact of research', *Research Global* February, 2007, viewed 20 February 2007, <http://www.atn.edu.au/docs/Research%20Global%20-%20Measuring%20the%20impact%20of%20research.pdf>

Elsevier 2006, 'Scirus adds new institutional repositories to Digital Archives', *Science Direct Info*, July/August 2006, viewed 28 September 2006, <http://info.sciencedirect.com/news/newsletters/connect/index.asp#3>.

Figueredo, E 2006, 'The numerical equivalence between the impact factor of journals and the quality of the articles', *Journal of the American Society for Information Science and Technology* v. 57, no. 11, p 1561, September.

Viewed 3 October, 2006, <http://www3.interscience.wiley.com. /cgi-bin/fulltext/112626123/PDFSTART>

Fingerman, S 2006, 'Web of Science and Scopus: current features and capabilities', Issues in Science and Technology Librarianship, Fall 2006. Viewed 8 December 2006. <http://www.istl.org/06-fall/electronic2.html>.

Houghton, John, Steele, Colin and Sheehan, Peter 2006, Research Communication Costs in Australia: Emerging Opportunities and Benefits: a Report to the department of Education, Science and Training. Canberra, DEST, viewed 5 October 2006, http://www.dest.gov.au/NR/rdonlyres/0ACB271F-EA7D-4FAF-B3F7-0381F441B175/13935/DEST_Research_Communications_Cost_Report_Sept2006.pdf.

Kisjes, Iris 2006, 'Scopus takes research evaluation where it's never gone before'. *Using Scopus for Bibliometric Analysis: a Practical Guide, Elsevier, viewed 19 December 2006, <http://www.elsevier.com/framework_librarians/LibraryConnect/LCP09/LCP09.pdf>*

Jacso, P 2006, 'Open access to scholarly full-text documents', *Online Information Review*, v. 30, no. 5, pp. 587-594.

Martello, A 2005, *Selection of content for the Web Citation Index: Institutional Repositories and subject-specific archives*, viewed 28 September 2006, <http://scientific.thomson.com/free/essays/selectionofmaterial/wci-selection/>

Meho, L. I. and Yang, K 2006, 'Multi-faceted approach to citation-based quality assessment for knowledge management', *Libraries: Dynamic Engines for the Knowledge and Information Society: IFLA General Conference and Council*, 72nd, 20-24 August 2006, Seoul, viewed 29 September 2006, <http://www.ifla.org/IV/ifla72/papers/146-Meho_Yang-en.pdf>.

National Health and Medical Research Council (Australia) 2006, *Project Grants Funding Policy for Funding Commencing in 2008*, viewed 12 December 2006, <http://www.nhmrc.gov.au/publications/_files/profundingpol.pdf>.

Pringle, J 2005, 'Partnering helps institutional repositories thrive', *KnowledgeLink Newsletter*, February, 2005, viewed 28 September 2006, <http://scientific.thomson.com/news/newsletter/2005-02/8264025/>.

Quint, B 2006, 'Institutional Repositories on Target: ARL Survey and Scopus/Scirus Features, *Information Today*, October 5, viewed 5 October 2006, <http://www.infotoday.com/newsbreaks/nb060828-2.shtml>.

Richardson, J 2006, 'Research Quality Framework as a catalyst for open access', AusWeb06: the Twelfth Australasian World Wide Web Conference,

Noosa Lakes, Qld, 1-5 July 2006, viewed 18 October 2006, <http://ausweb.scu.edu.au/aw06/papers/refereed/richardson/paper.html>

Rowlands I and Olivieri, R 2006, 'Overcoming barriers to research productivity: A case study in immunology and microbiology' *Publishing Research Consortium Summary Paper 1, viewed 28 September 2006,* <http://www.publishingresearch.org.uk/prcweb/PRCWeb.nsf/>.

Scopus 2006a, *Scopus Empowers Researchers with new Citation Tracker,* viewed 28 September 2006, <http://www.info.scopus.com/news/press/pr_060127.asp>.

Scopus 2006b, *Scopus Announces first-of-its-kind Customized Institutional Resources and Digital Archive Searches.* Viewed 28 September 2006, <http://info.scopus.com/news/press/pr_060821.asp>.

Scopus 2006c. 'Scopus Launches PatentCites & WebCites' *Managing Information*, September 2006, viewed 28 September 2006, http://www.managinginformation.com/news/content_show_full.php?id=5237

Scopus Content Coverage, May 2006, viewed 28 September 2006, <http://www.info.scopus.com/docs/content_coverage.pdf>.

Steele, C 2006a, 'Research with purpose', *The Australian*, 7 June 2006, p.42.

Steele, C, Butler, L and Kingsley, D 2006b. 'The Publishing imperative: the pervasive influence of publication metrics'. *Learned Publishing,* v. 19(4) pp. 277-290.

Swan, A 2006, 'Open Access - What has been going on?', viewed 6 November 2006, *HEP Libraries Webzine*, issue13, October 2006, <http://library.cern.ch/HEPLW/13/papers/1/>.

Thomson Scientific 2005, 'Thomson Scientific launches Web Citation Index: New scholarly index directs researchers to valuable content from repositories', viewed 28 September 2006, <http://scientific.thomson.com/press/2005/8298416/>.

Walter, G, and others 2003, 'Counting on citations: a flawed way to measure quality', *Medical Journal of Australia*, v. 178, issue 6, pp 280-281, viewed 28 September 2006, <http://www.mja.com.au/public/issues/178_06_170303/wal10537_fm.html>.