

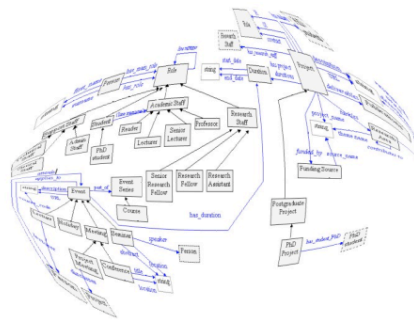
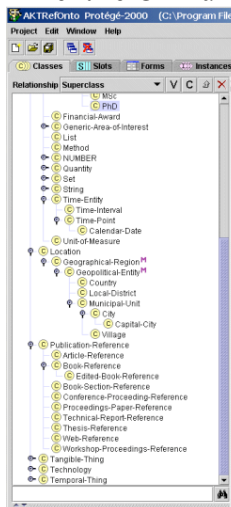
To appear in: In: Jacobs, N., Eds. Open Access: Key Strategic, Technical and Economic Aspects. Chandos. Note, final versions of figures and captions still to come, but their content and positions will be as shown here. The reference list will also be pruned, to cite or remove uncited works

The Open Research Web: A Preview of the Optimal and the Inevitable

Nigel Shadbolt, Tim Brody, Les Carr, and Stevan Harnad
University of Southampton

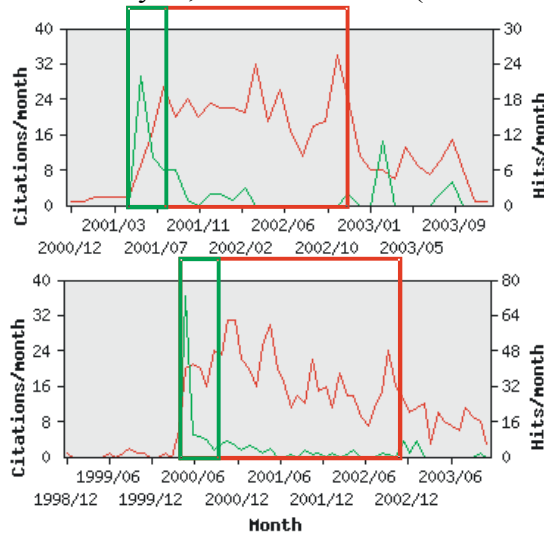
Most of this book has been about the past and the present of Open Access (OA). Let's now take a brief glimpse at its future, for it is already within reach and almost within sight: Once the optimal and inevitable outcome for the research literature has become actual, namely, all 2.5 million of the annual articles appearing in the planet's 24,000 peer-reviewed research journals are freely accessible online to all would-be users, then:

- (1) All their OAI metadata and full-texts will be harvested and reverse-indexed by services such as Google, OAIster (and new OAI/OA services), making it possible to search all and only the research literature in all disciplines using Boolean full-text search (and, or not, etc.).
- (2) Boolean full-text search will be augmented by Artificial Intelligence (AI) based text-analysis and classification techniques superior to human pre-classification, infinitely less time-consuming, and applied automatically to the entire OA full-text corpus.

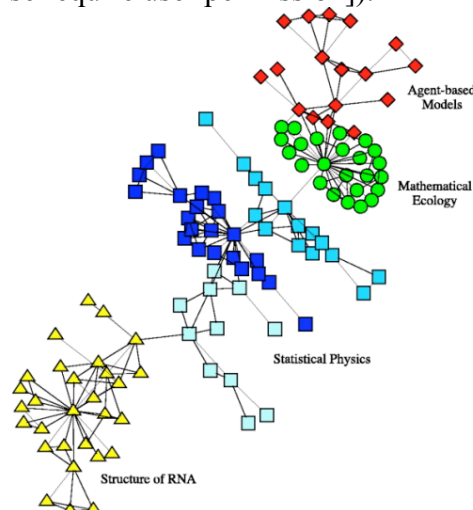


- (3) Articles and portions of articles will also be classified, tagged and annotated in terms of “ontologies” (lists of the kinds of things of interest in a subject domain, their characteristics, and their relations to other things) as provided by authors, users, other authorities, or automatic AI techniques, creating the OA research subset of the “semantic web” (Berners-Lee et al. 2001).
- (4) The OA corpus will be fully citation interlinked – every article forward-linked to every article it cites and backward-linked to every article that cites it -- making it possible to navigate all and only the research journal literature in all disciplines via citation-surfing instead of just ordinary link-surfing.
- (5) A CiteRank analog of Google’s PageRank algorithm will allow hits to be rank-ordered by weighted citation counts instead of just ordinary links (not all citations are equal: a citation by a much-cited author/article weighs more than a citation by a little-cited author/article; Page et al. 1999).

- (6) In addition to ranking hits by author/article/topic citation counts, it will also be possible to rank them by author/article/topic download counts (consolidated from multiple sites, caches, mirrors, versions) (Bollen et al. 2005; Moed 2005b).
- (7) Ranking and download/citation counts will not just be usable for searching but also (by individuals and institutions) for prediction, evaluation and other forms of analysis, on- and off-line (Moed 2005a).



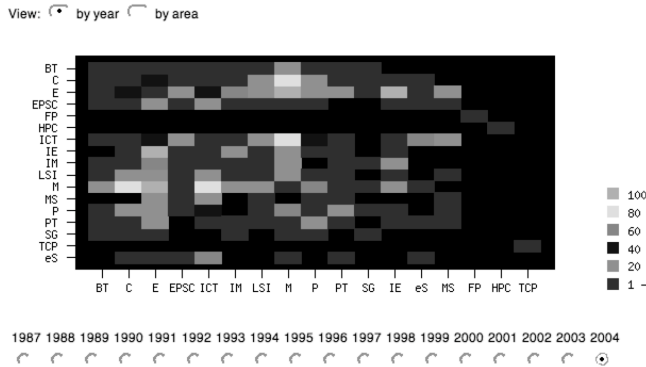
- (8) Correlations between earlier download counts and later citation counts will be available online, and usable for extrapolation, prediction and eventually even evaluation (Brody et al. 2005).
- (9) Searching, analysis, prediction and evaluation will also be augmented by co-citation analysis (who/what co-cited or was co-cited by whom/what?), co-authorship analysis, and eventually also co-download analysis (who/what co-downloaded or was co-downloaded by whom/what? [user identification will of course require user permission]).



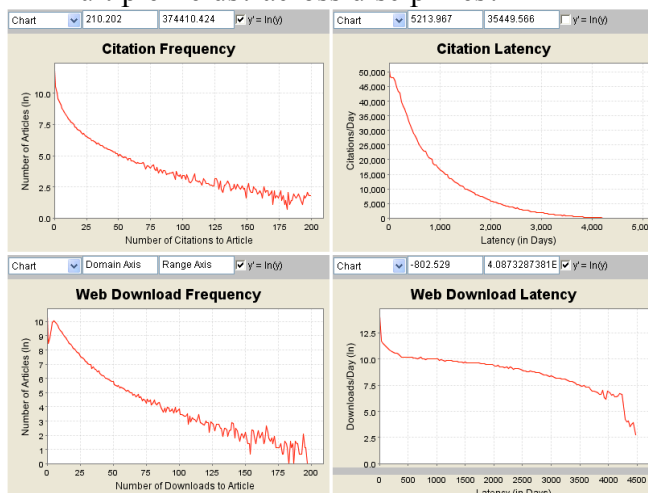
- (10) Co-text analysis (with AI techniques, including latent semantic analysis [what text and text-patterns co-occur with what? Landauer et al. 1998], semantic web analysis, and other forms of “semiometrics”; MacRae & Shadbolt 2006) will complement online and off-line citation, co-citation, download and co-download analysis (what texts have similar or related content or topics or users?).

- (11) Time-based (chronometric) analyses will be used to extrapolate early download, citation, co-download and co-citation trends, as well as correlations between downloads and citations, to predict research impact, research direction and research influences.

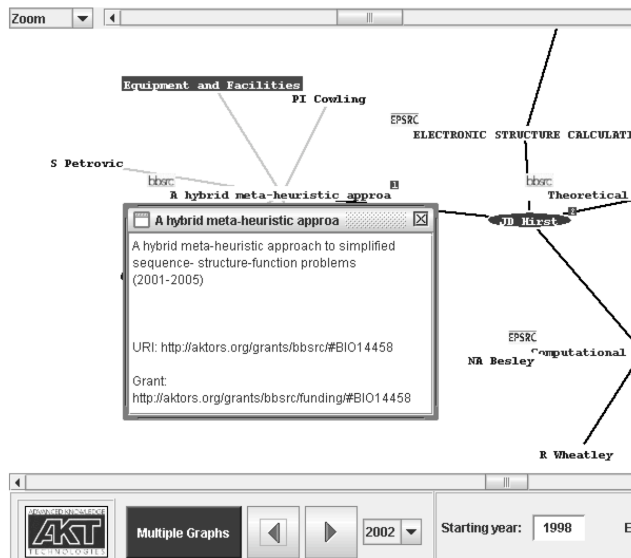
Collaboration over time



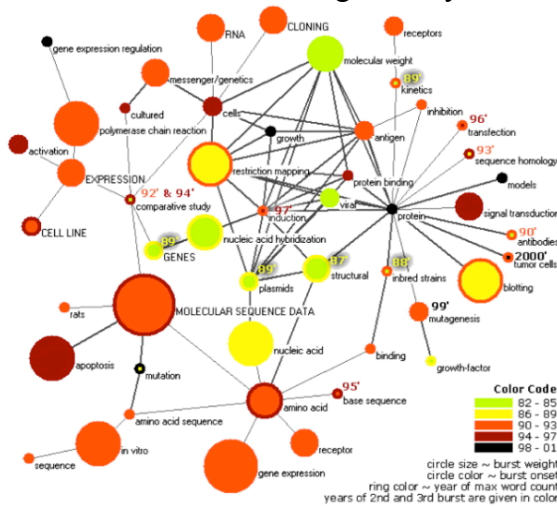
- (12) Authors, articles, journals, institutions and topics will also have “endogamy/exogamy” scores: how much do they cite themselves? in-cite within the same “family” cluster? out-cite across an entire field? across multiple fields? across disciplines?



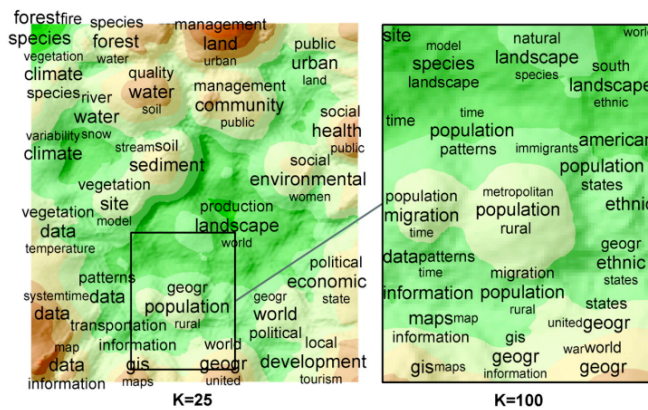
- (13) Authors, articles, journals, institutions and topics will also have latency and longevity scores for both downloads and citations: how quickly do citations/downloads grow? how long before they peak? how long-lived are they?
- (14) “Hub/authority” analysis will make it easier to do literature reviews, identifying review articles citing many articles (“hubs”) or key articles/authors (“authorities”) cited by many articles.
- (15) “Silent” or “unsung” authors or articles, uncited but important influences, will be identified (and credited) by co-citation and co-text analysis and through interpolation and extrapolation of semantic lines of influence.
- (16) Similarly, generic terms that are implicit in ontologies (but so basic that they are not explicitly tagged by anyone) -- as well as other “silent” influences, intermediating effects, trends and turning points -- can be discovered, extracted, interpolated and extrapolated from the patterns among the explicit properties such as citations and co-authorships, explicitly tagged features and relationships, and latent semantics.



- (17) Author names, institutions, URLs, addresses and email addresses will also be linked and disambiguated by this kind of triangulation.



- (18) Resource Description Framework (RDF) graphs (who is related to what, how?) will link objects in domain “ontologies”. For example, Social Network Analyses (SNA) on co-authors will be extended to other important relations and influences (projects directed, PhD students supervised etc.)
- (19) Co-text and semantic analysis will identify plagiarism as well as unnoticed parallelism and potential convergence.
- (20) A “degree-of-content-overlap” metric will be calculable between any two articles, authors, groups or topics.
- (21) Co-authorship, co-citation/co-download, co-text and chronometric path analyses will allow a composite “heritability” analysis of individual articles, indexing the amount and source of their inherited content, their original contribution, their lineage, and their likely future direction.



- (22) Cluster analyses and chronograms will allow connections and trajectories to be visualized, analyzed and navigated iconically.
- (23) User-generated tagging services (allowing users to both classify and evaluate articles they have used by adding tags anarchically) will complement systematic citation-based ranking and evaluation and author-based, AI-based, or authority-based semantic-web tagging, both at the article/author level and at the level of specific points in the text.
- (24) Commentaries -- peer-reviewed, moderated, and unmoderated -- will be linked to and from their target articles, forming a special, amplified class of annotated tags.
- (25) Referee-selection (for the peer reviewing of both articles and research proposals) will be greatly facilitated by the availability of the full citation-interlinked, semantically tagged corpus.
- (26) Deposit date-stamping will allow priority to be established.
- (27) Research articles will be linked to tagged research data, allowing independent re-analysis and replication.
- (28) The Research Web will facilitate much richer and more diverse and distributed collaborations, across institutions, nations, languages and disciplines (e-science, collaboratories).

Many of these future powers of the Open Access Research Web revolve around research impact: predicting it, measuring it, tracing it, navigating it, evaluating it, enhancing it. What is research impact?

Research Impact. The reason the employers and funders of scholarly and scientific researchers mandate that they should publish their findings (“publish or perish”) is that if research findings are kept in a desk drawer instead of being published then the research may as well not have been done at all. The *impact* of a piece of research is the degree to which it has been useful to other researchers and users in generating further research and applications: how much the work has been read, used, built-upon, applied and cited in other research as well as in educational, technological, cultural, social and practical applications (Moed 2005a).

The first approximation to a metric of research impact is the *publication* itself. Research that has not yielded any publishable findings has no impact. A second approximation metric of research impact is *where* it is published: To be accepted for publication, a research report must first be *peer-reviewed*, i.e., evaluated by qualified specialists who advise a journal editor on whether or not the paper can potentially

meet that journal's quality standards, and what revision needs to be done to make it do so. There is a hierarchy of journals in most fields, the top ones exercising the greatest selectivity, with the highest quality standards. So the second approximation impact metric for a research paper is the level in the journal quality hierarchy of the journal that accepts it. But even if published in a high-quality journal, a paper that no one goes on to read has no impact. So a third approximation impact metric comes from a paper's *usage* level. This was hard to calculate in print days, but in the online era, *downloads* can be counted (Kurtz et al 2002, 2004; Harnad & Brody 2004; Brody et al. 2005; Bollen et al. 2005; Moed 2005b). Yet even if a paper is downloaded and read, it may not be used -- not taken up, applied and built upon in further research and applications. The fourth metric and currently the closest approximation to a paper's research impact is accordingly whether it is not only published and read, but *cited*, which indicates that it has been used (by users other than the original author), as an acknowledged building block in further published work.

Being cited does not guarantee that a piece of work was important, influential and useful, and some papers are no doubt cited only to discredit them; but, on average, the more a work is cited, the more likely that it has indeed been used and useful (Garfield 1955, 1973, 1998, 1999; Brookes 1980/81; Wolfram 2003). Other estimates of the importance and productivity of research have proved to be correlated with its citation frequency. For example, every four years for two decades now, the UK Research Assessment Exercise (RAE) has been evaluating the research output of every department of every UK university, assigning each a rank along a 5-point scale on the basis of many different performance indicators, some consisting of peer judgments of the quality of published work, some consisting of objective metrics (such as prior research grant income, number of research students, etc.). A panel decides each department's rank and then each is funded proportionately. In many fields the ranking turns out to be most highly correlated with prior grant income, but it is almost as highly correlated with another metric: the total citation counts of each department's research output (Eysenck & Smith 2002, Harnad 2003; Harnad et al. 2003) *even though citations – unlike grant income -- are not counted explicitly in the RAE evaluation*. Because of the high correlation of the overall RAE outcome with metrics, it has now been decided, two decades after the inception of the RAE, to drop the peer-re-evaluation component entirely, and to rely entirely on research impact metrics (UK Office of Science and Technology 2006).

Measuring and Monitoring Article, Author and Group Research Impact.

ISI first provided the means of counting citations for articles, authors, or groups (see Garfield citations). We have used the same method – of linking citing articles to cited articles via their reference lists – to create Citebase <http://citebase.eprints.org/> (Brody 2003, 2004; Hitchcock et al 2003), a search engine like google, but based on citation links rather than arbitrary hyperlinks, and derived from the OA database instead of the ISI database. Citebase already embodies a number of the futuristic features we listed earlier. It currently ranks articles and authors by citation impact, co-citation impact or download impact and can be extended to incorporate multiple online measures (metrics) of research impact.

With only 15% of journal articles being spontaneously self-archived overall today, this is still too sparse a database to test and analyze the power of a scientometric engine like Citebase, but %OA is near 100% in a few areas of physics <http://arxiv.org>,

and this is where Citebase has been focused. Boolean search query results (using content words plus “and,” “or,” “not” etc.) can currently be quantified by Citebase and ranked in terms of article or author *download counts*, article/author *citation counts*, article/author *co-citedness counts* (how often is a sample of articles co-cited with – or by -- a given article or author?), *hub/authority counts* (an article is an “authority” the more it is cited by other authorities; this is similar to google’s PageRank algorithm, which does not count each of a cited article’s incoming citations as equal, but weights them by the citation counts of the article that is doing the citing; an article is a “hub” the more it cites authorities; Page et al. 1999). Citebase also has a *download/citation correlator* <http://www.citebase.org/analysis/correlation.php> which correlates downloads and citations across an adjustable time window. Natural future extensions of these metrics include *download growth-rate*, *latency-to-peak*, and *longevity indices* and *citation growth-rate*, *latency-to-peak*, and *longevity indices*.

So far, these metrics are only being used to rank-order the results of Citebase searches, as google is used. But they have the power to do a great deal more, and will gain still more power as %OA approaches 100%. The citation and download counts can be used to compare research impact, ranking articles, authors or groups; they can also be used to compare an individual’s own research impact with itself across time. The download and citation counts have also been found to be positively correlated with one another, so that early downloads, within 6 months of publication, can predict citations after 18 months or more (Brody et al. 2005). This opens up the possibility of time-series analyses, not only on articles’, authors’ or groups’ impact trajectories over time, but the impact trajectories of entire lines of research, when the citation/download analysis is augmented by *similarity/relatedness scores* derived from semantic analysis of text, for example, word and pattern co-occurrence, as in latent semantic analysis (Landauer et al 1998).

The natural objective is to develop a scientometric multiple regression equation for analyzing research performance and predicting research direction based on an OA database, beginning with the existing metrics. Such an equation of course needs to be validated against other metrics. The fourteen candidate predictors so far -- [1-4] article/author *citation counts*, *growth rates*, *peak latencies*, *longevity*; [5-8] the same metrics for downloads; [ix] download/citation correlation-based predicted citations; [10-12] hub/authority scores; [12-13] co-citation (with and by) scores; [14] co-text scores) -- can be made available open-endedly via tools like citebase, so that apart from users using them to rank search query results for navigation, individuals and institutions can begin using them to rank articles, authors or groups, validating them against whatever metrics they are currently using, or simply testing them open-endedly.

The method is essentially the same for navigation as well as analysis and evaluation: A search output -- or an otherwise selected set of candidates for ranking and analysis -- could each have the potential regression scores, whose weights could be set to 0 or a range from minimum to maximum, with an adjustable weight scale for each, normalizing to one across all the non-zero weights used. Students and researchers could use such an experimental battery of metrics as different ways of ranking literature search results; editors could use them for ranking potential referees; peer-reviewers could use them to rank the relevance of references; research assessors could use them to rank institutions, departments or research groups; institutional performance evaluators could use them to rank staff for annual review; hiring committees could use them to rank candidates; authors could use them to rank themselves against their competition.

It is important to stress that at this point all of this would not only be an unvalidated regression equation, to be used only experimentally, but that even after being validated against an external criterion or criteria, it would still need to be used in conjunction with human evaluation and judgment, and the regression weights would no doubt have to be set differently for different purposes, and always open for tweaking and updating. But it will begin ushering in the era of online, interactive scientometrics based on an Open Access corpus and in the hands of all users.

The software we have already developed and will develop, together with the growing webwide database of OA articles, and the data we will collect and analyse from it, will allow us to do several things for which the unique historic moment has arrived: (1) motivate more researchers to provide OA by self-archiving; (2) map the growth of OA across disciplines, countries and languages; (3) navigate the OA literature using citation-linking and impact ranking; (4) measure, extrapolate and predict the research impact of individuals, groups, institutions, disciplines, languages and countries; (5) measure research performance and productivity, (6) assess candidates for research funding; (7) assess the outcome of research funding, (8) map the course of prior research lines, in terms of individuals, institutions, journals, fields, nations; (9) analyze and predict the direction of current and future research trajectories; (10) provide teaching and learning resources that guide students (via impact navigation) through the large and growing OA research literature in a way that navigating the web via google alone cannot come close to doing.

At the forefront in the critical developments in OA across the past decade, our research team at Southampton University, UK:

- (i) hosts one of the first OA journals <http://psycprints.ecs.soton.ac.uk/> (since 1994),
- (ii) hosts the first journal OA preprint archive <http://www.bbsonline.org/> (since 1994)
- (iii) formulated the first OA self-archiving proposal (Okerson & O'Donnell 1995)
- (iv) founded one of the first central OA Archives <http://cogprints.org/> (1997)
- (v) founded the American Scientist Open Access Forum <http://amsci-forum.amsci.org/archives/American-Scientist-Open-Access-Forum.html> (1998)
- (vi) created the first (and now the most widely used) institutional OAI-compliant Archive-creating software <http://www.eprints.org/> (Sponsler & Van de Velde 2001, adopted by over 150 universities worldwide)
- (vii) co-drafted the BOAI (Budapest Open Access Initiative) self-archiving FAQ <http://www.eprints.org/self-faq/> (2001)
- (viii) created the first citation impact-measuring search engine <http://citebase.eprints.org/> (Hitchcock et al. 2003)
- (ix) created the first citation-seeking tool (to trawl the web for the full text of a cited reference) <http://paracite.eprints.org/> (2002)
- (x) designed the first OAI standardized CV http://paracite.eprints.org/cgi-bin/rae_front.cgi (2002)
- (xi) designed the first demonstration tool for predicting later citation impact from earlier download impact <http://citebase.eprints.org/analysis/correlation.php> (Brody et al. 2005)
- (xii) compiled the BOAI (Budapest Open Access Initiative) Eprints software Handbook <http://software.eprints.org/handbook/> (2003)
- (xiii) formulated the model self-archiving policy for departments and institutions <http://software.eprints.org/handbook/departments.php> (2003)
- (xiv) created and maintain ROAR, the Registry of Open Access Repositories worldwide <http://archives.eprints.org> (2003)

- (xv) collaborated in the creation and maintenance of the ROMEEO directory of journals' self-archiving policies <http://romeo.eprints.org/> (2004: of the top 9,000 journals across all fields, 92% already endorse author self-archiving)
- (xvi) created and maintain ROARMAP, a registry of institutions' self-archiving policies <http://www.eprints.org/signup/sign.php> (2004)
- (xvii) piloted the paradigm of collecting, analyzing and disseminating data on the magnitude of the OA impact advantage and the growth of OA across all disciplines worldwide http://citebase.eprints.org/isi_study/ (Harnad et al. 2004)

The multiple online research impact metrics we are developing will allow the rich new database, the Research Web, to be navigated, analyzed, mined and evaluated in powerful new ways that were not even conceivable in the paper era – nor even in the online era, until the database and the tools became openly accessible for online use by all: by researchers, research institutions, research funders, teachers, students, and even by the general public that funds the research and for whose benefit it is being conducted: Which research is being used most? By whom? Which research is growing most quickly? In what direction? under whose influence? Which research is showing immediate short-term usefulness, which shows delayed, longer term usefulness, and which has sustained long-lasting impact? Which research and researchers are the most authoritative? Whose research is most using this authoritative research, and whose research is the authoritative research using? Which are the best pointers (“hubs”) to the authoritative research? Is there any way to predict what research will have later citation impact (based on its earlier download impact), so junior researchers can be given resources before their work has had a chance to make itself felt through citations? Can research trends and directions be predicted from the online database? Can text content be used to find and compare related research, for influence, overlap, direction? Can a layman, unfamiliar with the specialized content of a field, be guided to the most relevant and important work? These are just a sample of the new online-age questions that the Open Research Web will begin to answer.

References

- Adams, J. (2005) Early citation counts correlate with accumulated impact. *Scientometrics*, 63 (3): 567-581
- Ajiferuke, I. , & Wolfram, D. (2004). Modeling the characteristics of web page outlinks. *Scientometrics*, 59(1), 43-62.
- Alani, H., Nicholas, G., Glaser, H., Harris, S. and Shadbolt, N. (2005) Monitoring Research Collaborations Using Semantic Web Technologies. 2nd European Semantic Web Conference (ESWC).
- Antelman, K. (2004) Do Open-Access Articles Have a Greater Research Impact? *College and Research Libraries*, 65(5):372-382
- Baudoin, L. , Haefner-Cavaillon, N. , Pinhas, N. , Mouchet, S. and Kordon, C. (2004) Bibliometric indicators: realities, myth and prospective *Med Sci (Paris)*, 20 (10):909-15
- Bauer, K. and Bakkalbasi, N. (2005) An Examination of Citation Counts in a New Scholarly Communication Environment *D-Lib Magazine*, 11(9)
- Bauer, K. and Bakkalbasi, N. (2005) An Examination of Citation Counts in a New Scholarly Communication Environment *D-Lib Magazine*, 11(9), September 2005.
- Berners-Lee, T, Hendler, J. and Lassila, O. (2001) The Semantic Web, *Scientific American* 284 (5): 34-43.
<http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>
- Belew, R. (2005) Scientific impact quantity and quality: Analysis of two sources of bibliographic data *Arxiv. org*, cs. IR/0504036
- Berners-Lee, T., De Roure, D., Harnad, S. and Shadbolt, N. (2005) Journal publishing and author self-archiving: Peaceful Co-Existence and Fruitful Collaboration.
- Open Letter to Research Councils UK.
- Bollen, J. and Luce, R. (2002) Evaluation of Digital Library Impact and User Communities by Analysis of Usage Patterns *D-Lib Magazine*, Vol. 8, No. 6
- Bollen, J. , Van de Sompel, H. , Smith, J. and Luce, R. (2005) Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing and Management*, 41(6): 1419-1440,
- Bollen, J. , Vemulapalli, S. S. , Xu, W. and Luce, R. (2003) Usage Analysis for the Identification of Research Trends in Digital Libraries *D-Lib Magazine*, Vol. 9, No. 5.
- Borgman, C. L. and Furner, J. (2002) Scholarly Communication and Bibliometrics , *Annual Review of Information Science and Technology*, Vol. 36, edited by B. Cronin, 2002
- Brody, T. (2003) Citebase Search: Autonomous Citation Database for e-print Archives, sinn03 *Conference on Worldwide Coherent Workforce, Satisfied Users - New Services For Scientific Information*, Oldenburg, Germany, September 2003
- Brody, T. (2004) Citation Analysis in the Open Access *World Interactive Media International*
- Brody, T. , Harnad, S. and Carr, L. (2005) Earlier Web Usage Statistics as Predictors of Later Citation Impact. *Journal of the American Association for Information Science and Technology (JASIST)*, in press).
- Brody, T. , Kampa, S. , Harnad, S. , Carr, L. and Hitchcock, S. (2003) Digitometric Services for Open Archives Environments. In *Proceedings of European Conference on Digital Libraries 2003*, pages pp. 207-220, Trondheim, Norway.

- Brookes, B. C. (1980/1981). The foundations of information science. *Journal of Information Science* 2 (1980) pp. 125-133, 209-221, 269-173; 3 (1981) pp. 3-12.
- Carr, L. and Harnad, S. (2005) *Keystroke Economy: A Study of the Time and Effort Involved in Self-Archiving*
- Carr, L. , Hitchcock, S. , Hall, W. and Harnad, S. (2000) A usage based analysis of CoRR. *ACM SIGDOC Journal of Computer Documentation* 24(2):pp. 54-59.
- Carr, L. , Hitchcock, S. , Hall, W. and Harnad, S. (2001) Enhancing OAI Metadata for Eprint Services: two proposals. In *Proceedings of Experimental OAI-based Digital Library Systems Workshop 5*, pages pp. 54-59, Darmstadt. Brody, T. , Jiao, Z. , Hitchcock, S. , Carr, L. and Harnad, S. , Eds.
- Cockerill, M. J. (2004) Delayed impact: ISI's citation tracking choices are keeping scientists in the dark *BMC Bioinformatics* 2004, 5:93
- Darmoni, S. J. et al. (2002) Reading factor: a new bibliometric criterion for managing digital libraries, *Journal of the Medical Library Association*, Vol. 90, No. 3, July 2002
- De Roure, D., Jennings, N. R. and Shadbolt, N. R. (2005) The Semantic Grid: Past, Present and Future. *Proceedings of the IEEE* 93(3) pp. 669-681
- Diamond, Jr. , A. M. (1986) What is a Citation Worth? *J. Hum. Resour.* , 21:200-15, 1986
- Donohue JM, Fox JB (2000) A multi-method evaluation of journals in the decision and management sciences by US academics. *Omega-International Journal of Mangement Science* 28 (1): 17-36
- Garfield, E. (1955) Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, Vol:122, No:3159, p. 108-111
- Garfield, E. (1973) Citation Frequency as a Measure of Research Activity and Performance, in *Essays of an Information Scientist*, 1: 406-408, 1962-73, Current Contents, 5
- Garfield, E. (1988) Can Researchers Bank on Citation Analysis? *Current Comments*, No. 44, October 31, 1988
- Garfield, E. (1998) The use of journal impact factors and citation analysis in the evaluation of science, *41st Annual Meeting of the Council of Biology Editors*, Salt Lake City, UT, May 4, 1998
- Garfield, E. (1999) Journal impact factor: a brief review *CMAJ*, 161 (8), October 19, 1999
- Hajjem, C., Gingras, Y., Brody, T., Carr, L. & Harnad, S. (2005) *Across Disciplines, Open Access Increases Citation Impact*. (manuscript in preparation).
- Hajjem, C. (2005) *Analyse de la variation de pourcentages d'articles en accès libre en fonction de taux de citations* <http://www.crsc.uqam.ca/lab/chawki/ch.htm>
- Harnad, S. (1990) Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry *Psychological Science* 1: 342 - 343 (reprinted in Current Contents 45: 9-13, November 11 1991).
- Harnad, S. (1991) Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. *Public-Access Computer Systems Review* 2 (1): 39 – 53
- Harnad, S. (1995) The Subversive Proposal. In: Ann Okerson & James O'Donnell (Eds.) *Scholarly Journals at the Crossroads; A Subversive Proposal for Electronic Publishing*. Washington, DC., Association of Research Libraries, June 1995.

- Harnad, Stevan (1998) For Whom the Gate Tolls? Free the Online-Only Refereed Literature. American Scientist Forum
- Harnad, S. (2001) Research Access, Impact and Assessment, *Times Higher Education Supplement*, 1487: p. 16. , 2001
- Harnad, S. (2003) Enhance UK research impact and assessment by making the RAE webmetric Author eprint, in *Times Higher Education Supplement*, 6 June 2003, p. 16
- Harnad, S. (2003) Maximizing university research impact through self-archiving *Jekyll.com*, No. 7, December 2003
- Harnad, S. (2004) The Green Road to Open Access: A Leveraged Transition *American Scientist Open Access Forum*, January 07, 2004
- Harnad, S. (2005) Maximising the Return on UK's Public Investment in Research (submitted)
- Harnad, S. (2005a) Canada is not Maximizing the Return on its Research Investment. *Research Money*, November 2005.
- Harnad, S. (2005b) Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold. *Ariadne* 42.
- Harnad, S. and Brody, T. (2004a) Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*, Vol. 10 No. 6
- Harnad, S. and Brody, T. (2004) Prior evidence that downloads predict citations. *British Medical Journal* online.
- Harnad, S. and Carr, L. (2000) Integrating, Navigating and Analyzing Eprint Archives Through Open Citation Linking (the OpCit Project). *Current Science* 79(5):pp. 629-638.
- Harnad, S. , Brody, T. , Vallieres, F. , Carr, L. , Hitchcock, S. , Gingras, Y. , Oppenheim, C. , Stamerjohanns, H. and Hilf, E. (2004) The Access/Impact Problem and the Green and Gold Roads to Open Access. *Serials Review*, Vol. 30, No. 4, 310-314
- Harnad, S. , Carr, L. and Brody, T. (2001) How and Why to Free All Refereed Research from Access- and Impact-Barriers, Now How and Why To Free All Refereed Research. *High Energy Physics Library Webzine* 4.
- Harnad, S. , Carr, L. , Brody, T. and Oppenheim, C. (2003) Mandated online RAE CVs linked to university eprint archives: Enhancing UK research impact and assessment *Ariadne*, issue 35, April 2003
- Hitchcock, S. , Brody, T. , Gutteridge, C. , Carr, L. and Harnad, S. (2003b) The Impact of OAI-based Search on Access to Research Journal Papers, *Serials*, Vol. 16, No. 3, November 2003, 255-260
- Hitchcock, S. , Brody, T. , Gutteridge, C. , Carr, L. , Hall, W. , Harnad, S. , Bergmark, D. and Lagoze, C. (2002) Open Citation Linking: The Way Forward. *D-Lib Magazine* 8(10).
- Hitchcock, S. , Carr, L. , Jiao, Z. , Bergmark, D. , Hall, W. , Lagoze, C. and Harnad, S. (2000) Developing services for open eprint archives: globalisation, integration and the impact of links. In *Proceedings of the 5th ACM Conference on Digital Libraries*, San Antonio, Texas, June 2000. , pages pp. 143-151.
- Hitchcock, S. , Woukeu, A. , Brody, T. , Carr, L. , Hall, W. and Harnad, S. (2003) Evaluating Citebase, an open access Web-based citation-ranked search and impact discovery service. *Technical Report ECSTR-IAM03-005*, School of Electronics and Computer Science, University of Southampton

Holmes, A. and Oppenheim, C. (2001) Use of citation analysis to predict the outcome of the 2001 Research Assessment Exercise for Unit of Assessment (UoA) 61: *Library and Information Management Information Research*, Vol. 6, No. 2, January 2001

Jacsó, P. (2004) *The Future of Citation Indexing - Interview with Dr. Eugene Garfield*.

Jacsó, P. (2005) Google Scholar (Redux), *Thomson Gale*, June 2005

Kurtz, M. J. (2004) Restrictive access policies cut readership of electronic research journal articles by a factor of two Presented at *National Policies on Open Access (OA) Provision for University Research Output*: an International meeting, Southampton, 19 February 2004

Jewell, M. O., Lawrence, K. F., Tuffield, M. M., Prugel-Bennett, A., Millard, D. E., Nixon, M. S., schraefel, m. c. and Shadbolt, N. R. (2005) OntoMedia: An Ontology for the Representation of Heterogeneous Media. In Proceedings of Multimedia Information Retrieval Workshop (in press), Brazil.

Kurtz, M. J. , Eichhorn, G. , Accomazzi, A. , Grant, C. , Demleitner, M. , Murray, S. S. , Martimbeau, N. and Elwell, B. (2003) The NASA Astrophysics Data System: Sociology, Bibliometrics, and Impact, *Journal of the American Society for Information Science and Technology*

Kurtz, M. J. , Eichhorn, G. , Accomazzi, A. , Grant, C. S. , Demleitner, M. , Murray, S. S. (2004) The Effect of Use and Access on Citations, *Information Processing and Management*, 41 (6): 1395-1402

Kurtz, M. J. , Eichhorn, G. , Accomazzi, A. , Grant, C. S. , Demleitner, M. , Murray, S. S. , Martimbeau, N. and Elwell, B. (2003a) The Bibliometric Properties of Article Readership Information, *Journal of the American Society for Information Science and Technology*, 56 (2): 111-128,

Kurtz, M. J. , Eichhorn, G. , Accomazzi, A. , Grant, C. S. , Thompson, D. M. , Bohlen, E. H. and Murray, S. S. (2002) The NASA Astrophysics Data System: Obsolescence of Reads and Cites. *Library and Information Services in Astronomy IV*, edited by B. Corbin, E. Bryson, and M. Wolf

Kurtz, M. J. , Eichhorn, G. , Accomazzi, A. , Grant, C. S. , Demleitner, M. and Murray, S. S. (2004a) Worldwide Use and Impact of the Nasa Astrophysics Data System Digital Library, *Journal of the American Society for Information Science and Technology*, Vol. 56, No. 1, 36-45

LaGuardia, C. (2005) Scopus vs. Web of Science, *Library Journal*, 130(1); 40, 42, January 15, 2005

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.

Lawrence, S. (2001) Free online availability substantially increases a paper's impact *Nature*, 31 May 2001

Lawrence, S. , Giles, C. L. , Bollacker, K. (1999), Digital Libraries and Autonomous Citation Indexing, *IEEE Computer*, Vol. 32, No. 6, 67-71, 1999

Lee KP, Schotland M, Bacchetti P, Bero LA (2002) Association of journal quality indicators with methodological quality of clinical research articles. *Journal of the American Medical Association (JAMA)* 287 (21): 2805-2808

Liu, X. , Brody, T. , Harnad, S. , Carr, L. , Maly, K. , Zubair, M. and Nelson, M. L. (2002) *A Scalable Architecture for Harvest-Based Digital Libraries - The ODU/Southampton Experiments*.

McRae-Spencer, D. M. & Shadbolt, N.R. (2006) Semiometrics: Producing a Compositional View of Influence. (preprint)

- McVeigh, M. E. (2004) Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation Patterns *Thomson Scientific*
- Millard, D. E. , Bailey, C. P. , Brody, T. D. , Dupplaw, D. P. , Hall, W. , Harris, S. W. , Page, K. R. , Power, G. and Weal, M. J. (2003) Hyperdoc: An Adaptive Narrative System for Dynamic Multimedia Presentations. *Technical Report ECSTR-IAM02-006*, Electronics and Computer Science, University of Southampton.
- Moed, H. F. (2005a) *Citation Analysis in Research Evaluation*. NY Springer.
- Moed, H. F. (2005b) Statistical Relationships Between Downloads and Citations at the Level of Individual Documents Within a Single Journal, *Journal of the American Society for Information Science and Technology*, 56(10): 1088-1097
- Odlyzko, A. M. (1995) Tragic loss or good riddance? The impending demise of traditional scholarly journals, *Intern. J. Human-Computer Studies* 42 (1995), pp. 71-122
- Odlyzko, A. M. (2002) The rapid evolution of scholarly communication *Learned Publishing*, 15(1), 7-19
- Page, L., Brin, S., Motwani, R., Winograd, T. (1999) The PageRank Citation Ranking: Bringing Order to the Web. <http://dbpubs.stanford.edu:8090/pub/1999-66>
- Perneger, T. V. (2004) Relation between online "hit counts" and subsequent citations: prospective study of research papers in the British Medical Journal. *British Medical Journal* 329:546-547
- Pringle, J. (2004) Do Open Access Journals have Impact? *Nature Web Focus: access to the literature*
- Ray J, Berkwits M, Davidoff F (2000) The fate of manuscripts rejected by a general medical journal. *American Journal of Medicine* 109 (2): 131-135.
- Schwarz, G. (2004) Demographic and Citation Trends in Astrophysical Journal Papers and Preprints *Bulletin of the American Astronomical Society*, accepted
- Seglen, P. O. (1997) Why the impact factor of journals should not be used for evaluating research, *British Medical Journal*, 314:497
- Shin E. J. (2003) Do Impact Factors change with a change of medium? A comparison of Impact Factors when publication is by paper and through parallel publishing. *Journal of Information Science*, 29 (6): 527-533, 2003
- Smith, A. and Eysenck, M. (2002) The correlation between RAE ratings and citation counts in psychology *Technical Report*, Psychology, Royal Holloway College, University of London, June 2002
- Spink, A. , Wolfram, D. & Jansen, B. J. & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*. 52(3), 226-234.
- Sponsler, E. & Van de Velde E. F. (2001) Eprints.org Software: A Review. *Sparc E-News*, August-September 2001.
- Suber, P. (2003) Removing the Barriers to Research: An Introduction to Open Access for Librarians, *College & Research Libraries News*, 64, February, 92-94, 113
- Swan, A., Needham, P., Proberts, S., Muir, A., Oppenheim, C., O'Brien, A., Hardy, R. and Rowland, F. (2005) Delivery, Management and Access Model for E-prints and Open Access Journals within Further and Higher Education. *Technical Report, JISC, HEFCE*.
- Swan, A., Needham, P., Proberts, S., Muir, A., Oppenheim, C., O'Brien, A., Hardy, R., Rowland, F. and Brown, S. (2005) Developing a model for e-prints and open access journal content in UK further and higher education. *Learned Publishing* 18(1):pp. 25-40.

- Swan, A. and Brown, S. (2004) Authors and open access publishing. *Learned Publishing* 17(3):pp. 219-224.
- Tenopir, Carol and Donald W. King. *Towards Electronic Journals: Realities for Scientists, Librarians and Publishers*. 2000. Washington, D. C. SLA Publishing.
- Testa, J. and McVeigh, M. E. (2004) The Impact of Open Access Journals: A Citation Study from *Thomson ISI*
- UK Office of Science and Technology (2006) Science and innovation investment framework 2004-2014: next steps http://www.hm-treasury.gov.uk/media/1E1/5E/bud06_science_332.pdf
- Vaughan, L. and Shaw, D. (2004) Can Web Citations be a Measure of Impact? An Investigation of Journals in the Life Sciences. *Proceedings of the 67th ASIS&T Annual Meeting*, Vol. 41 pp. 516-526
- Walter, G. , Bloch, S. , Hunt, G. and Fisher, K. (2003) Counting on citations: a flawed way to measure quality *MJA*, 2003, 178 (6): 280-281
- Wolfram, D. (2003). *Applied informetrics for information retrieval research*. Westport, CT: Libraries Unlimited.
- Wouters, P. (1999) *The Citation Culture* PhD Thesis, University of Amsterdam
- Wren, J. D. (2005) Open access and openly accessible: a study of scientific publications shared via the internet. *British Medical Journal*, 330:1128,
- Yamazaki S (1995) Refereeing System of 29 Life-Science Journals Preferred by Japanese Scientists *Scientometrics* 33 (1): 123-129
- Youngen, G. (1998) Citation Patterns Of The Physics Preprint Literature With Special Emphasis On The Preprints Available Electronically ACRL/STS 6/29/97
- Youngen, G. K. (1998) Citation Patterns to Electronic Preprints in the Astronomy and Astrophysics Literature Library and Information Services in Astronomy III, ASP Conference Series, Vol. 153, 1998
- Zhao, D. (2005) Challenges of scholarly publications on the Web to the evaluation of science -- A comparison of author visibility on the Web and in print journals, *Information Processing and Management*, 41:6, 1403-1418