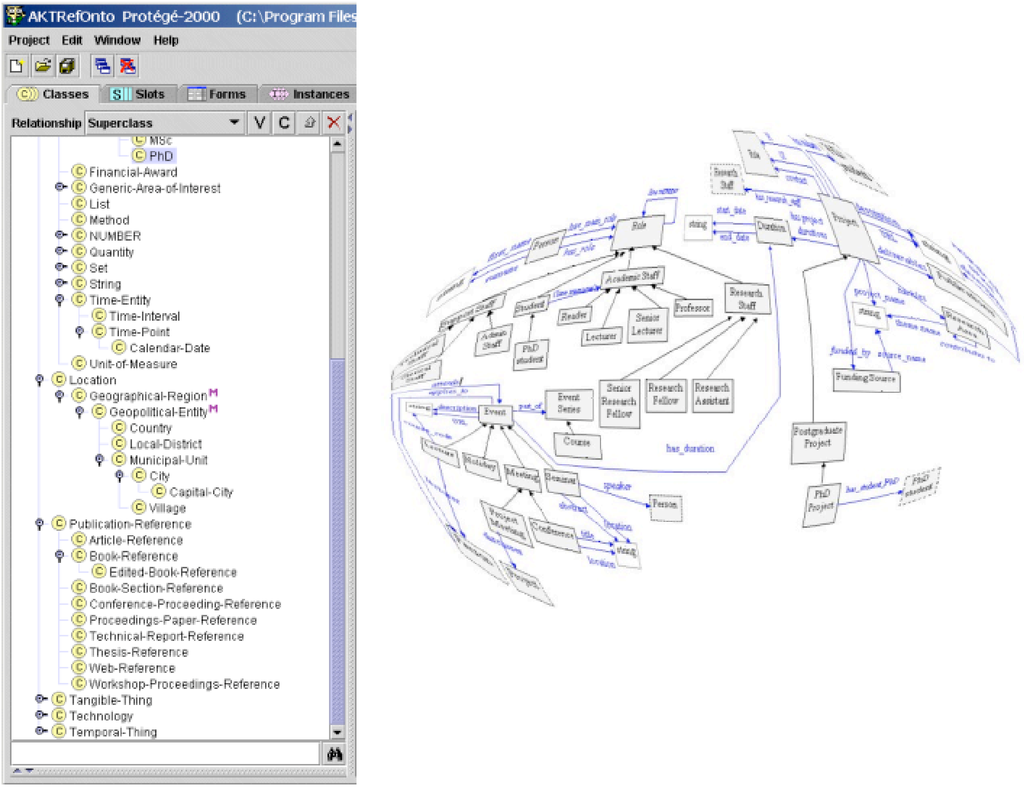
Chapter 21: The Open Research Web

Nigel Shadbolt, Tim Brody, Les Carr and Stevan Harnad

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. Imagine a world in which the optimal outcome for the research literature has become actual: With all 2.5 million of the annual articles in the planet's 24,000 peer-reviewed research journals freely accessible online to all would-be users (Odlyzko 1995; Okerson & O'Donnell 1995; Berners-Lee et al. 2005; DeRoure et al. 2005):

(1) All their OAI metadata and full-texts will be harvested, inverted and indexed by services such as *Google*, *OAIster* and still newer 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.



[Insert Figure 21.1]

Figure 21.1: Various visualisations of an ontology

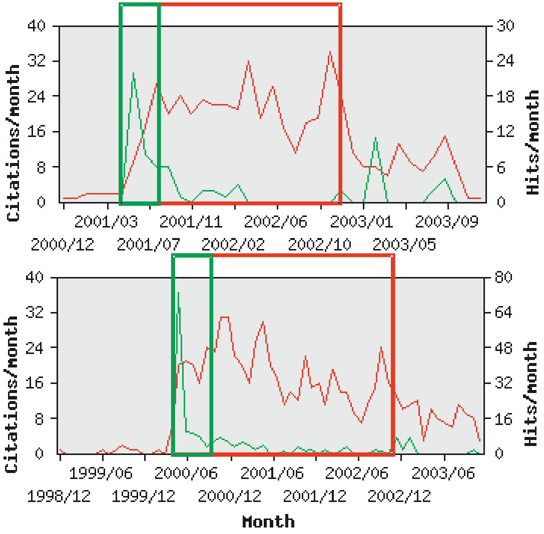
(3) Articles and portions of articles will also be classified, tagged and annotated in terms of "ontologies" (lists of the kinds of things of interest in a subject domain, their characteristics, and their relations to other things, see Figure 21.1.) as provided by authors, users, other authorities, or 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 analogue 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) (Adams 2005; 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).

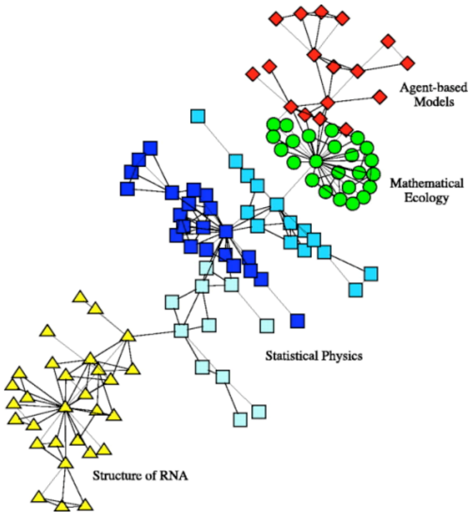


[Insert Figure 21.2]

Figure 21.1: An earlier window of downloads (green) may predict a later window of citations (red) (from Brody et al. 2006)

(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*, 2006).

(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]).

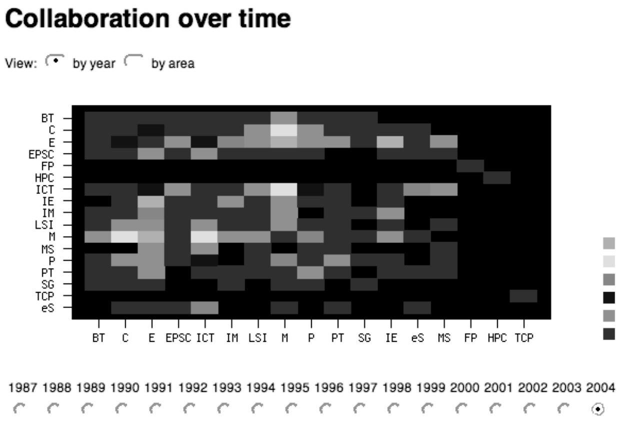


[Insert Figure 21.3]

Figure 21.3: A small co-authorship depicting collaborations between scientists across topic and subject boundaries (from Newman, 2004)

(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.

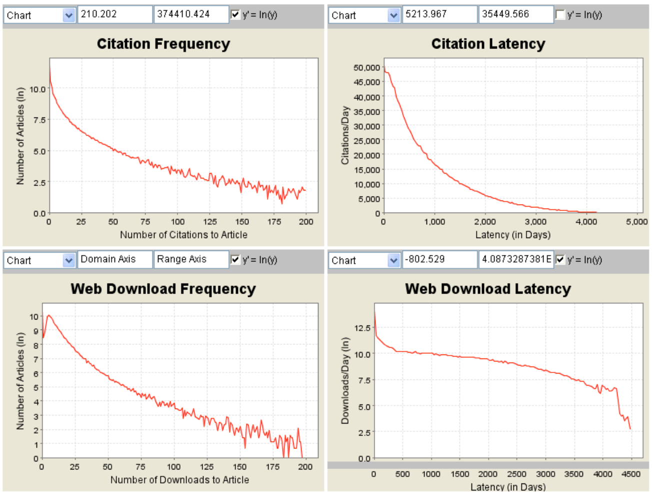

Collaboration over time

View: ● by year ○ by area

[Insert Figure 21.4]

Figure 21.4: Results of a simple chronometric analysis, showing collaboration via endogamy/exogamy scores (Alani et al 2005)

(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?



[Insert Figure 21.5]

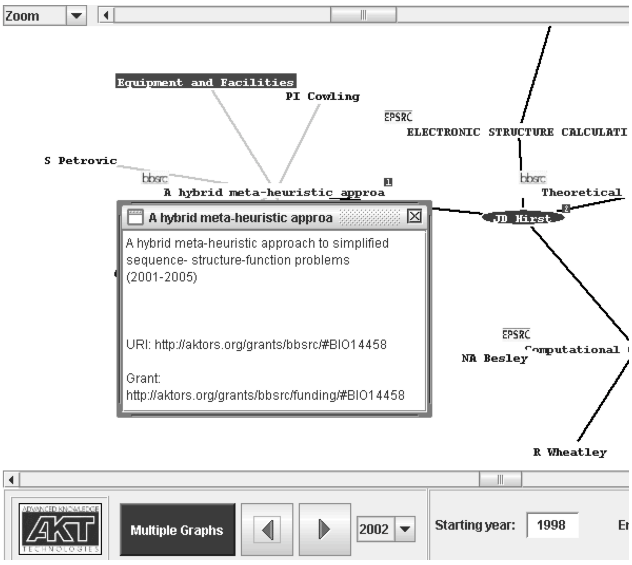
Figure 21.5: Time course of downloads and citations (Brody et al. 2006).

(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 (Klienber 1999(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.



A hybrid meta-heuristic approa

A hybrid meta-heuristic approach to simplified sequence- structure-function problems (2001-2005)

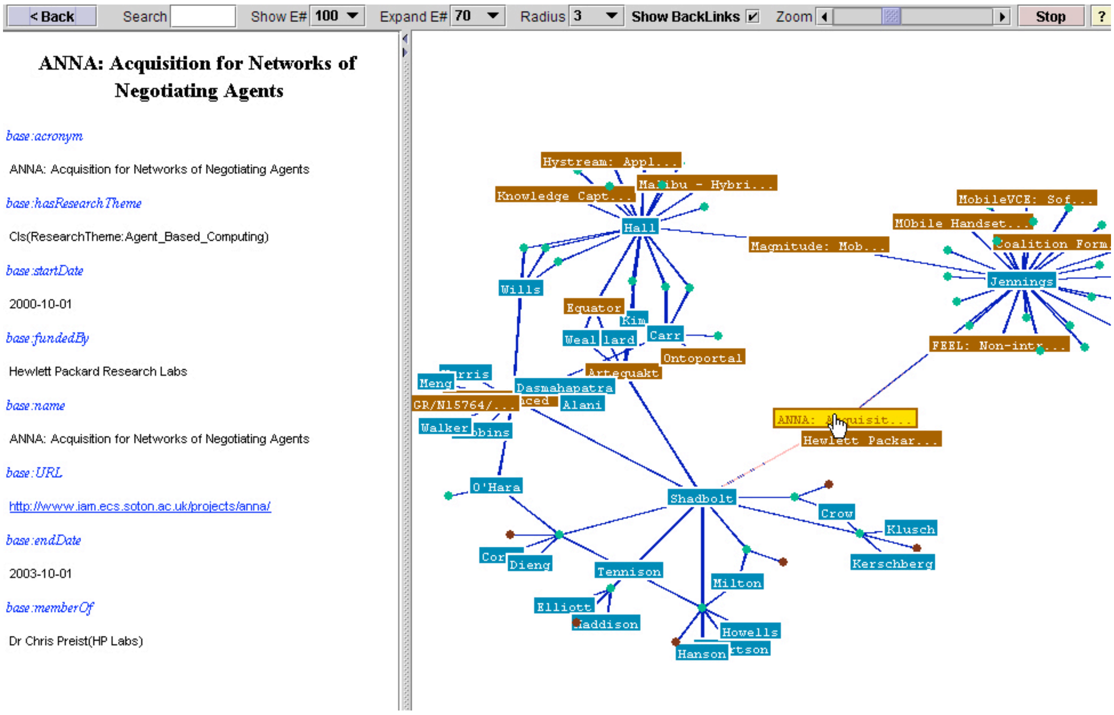
URI: http://aktors.org/grants/bbsrc/#BIO14458

Grant: http://aktors.org/grants/bbsrc/funding/#BIO14458

[Insert Figure 21.6]

Figure 21.6: Linked map of research entities

(17) Author names, institutions, projects, URLs, addresses and email addresses will also be linked and disambiguated by this kind or triangulation (Figure 21.6).



< Back   Search   Show E# 100   Expand E# 70   Radius 3   Show BackLinks ☑   Zoom   Stop   ?

ANNA: Acquisition for Networks of Negotiating Agents

base:acronym

ANNA: Acquisition for Networks of Negotiating Agents

base:hasResearchTheme

Cls(ResearchTheme:Agent_Based_Computing)

base:startDate

2000-10-01

base:fundedBy

Hewlett Packard Research Labs

base:name

ANNA: Acquisition for Networks of Negotiating Agents

base:URL

http://www.iam.ecs.soton.ac.uk/projects/anna/

base:endDate

2003-10-01

base:memberOf

Dr Chris Preist(HP Labs)

[Insert Figure 21.7]

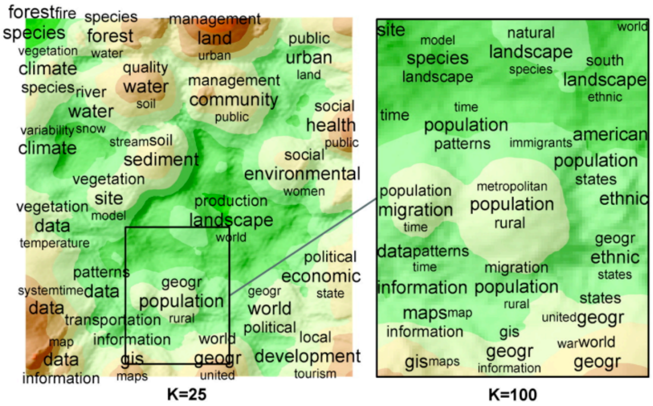
Figure 21.7: A Social Network Analysis Tool Rendering an RDF Graph (Alani et al 2003)

(18) Resource Description Framework (RDF) graphs (who is related to what, how?) will link objects in domain 'ontologies'. For example, Social Network Analyses 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.



[Insert Figure 21.8]

Figure 21.8: A self organising map supporting navigable visualisation of a research domain (from Skupin, 2004)

(22) Cluster analyses and chronograms will allow connections and trajectories to be visualised, analysed 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 (Connotea).
(24) Commentaries – peer-reviewed, moderated, and unmoderated – will be linked to and from their target articles, forming a special, amplified class of annotated tags (Harnad 1978, 1990).
(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*, that is, 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 2004; Harnad & Brody 2004; Brody et al. 2006; 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; Wolfram 2003). Other estimates of the importance and productivity of research have proved to be correlated with its citation frequency. For example, about every six 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, or number of research students). 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 (Smith & Eysenck 2002; 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, two decades after the inception of the RAE:

> "*the Government has a firm presumption that after the 2008 RAE the system for assessing research quality and allocating "quality-related" (QR) research funding to universities from the Department for Education and Skills will be mainly metrics-based*" (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 Search* (Brody 2003, 2004), 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 analyse the power of a scientometric engine like Citebase, but %OA is near 100% in a few areas of physics that use *arXiv*, and this is where Citebase has been focused. Boolean search query results (using content words plus 'and', 'or', 'not' and so on) 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 web links as equal, but weights them by the number of links to the linking page; an article is a 'hub' the more it cites authorities; Page et al. 1999). Citebase also has a *Citebase download/citation correlator*, 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 six months of publication, can predict citations after 18 months or more (Brody et al. 2006). 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).

Benefits

The natural objective is to develop a scientometric multiple regression equation for analysing 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, normalising 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) analyse 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, *Psycoloquy* (since1994);
(ii) hosts the first journal OA preprint archive, *BBSPrints* (since 1994);
(iii) formulated the first OA self-archiving proposal (Okerson & O'Donnell 1995);
(iv) founded one of the first central OA Archives, *Cogprints* (1997);
(v) founded the *American Scientist Open Access Forum* (1998);
(vi) created the first (and now the most widely used) institutional OAI-compliant archive-creating software, *Eprints* (Sponsler & Van de Velde 2001), adopted by over 150 universities worldwide;
(vii) co-drafted the Budapest Open Access Initiative, *BOAI self-archiving FAQ* (2001);
(viii) created the first citation impact-measuring search engine, *Citebase Search* (Hitchcock et al. 2003);
(ix) created the first citation-seeking tool (to trawl the web for the full text of a cited reference), Paracite (2002);
(x) designed the first OAI standardised CV, *Template for UK Standardized CV for Research Assessment* (2002);
(xi) designed the first demonstration tool for predicting later citation impact from earlier download impact, the *Citebase download / citation correlator* (Brody et al. 2006);
(xii) compiled the Budapest Open Access Initiative, *BOAI Eprints software Handbook* (2003);
(xiii) formulated the model self-archiving policy for departments and institutions, *Actions for Departments to Achieve Open Access* (2003);
(xiv) created and maintain ROAR, the *Registry of Open Access Repositories* worldwide (2003)
(xv) collaborated in the creation and maintenance of the ROMEO directory of journals' self-archiving policies, *Eprints Journal Policies* (2004: of the top 9,000 journals across all fields, 92% already endorse author self-archiving);
(xvi) created and maintain ROARMAP, the *Registry of Open Access Repository Material Archiving Policies* (2004);
(xvii) piloted the paradigm of collecting, analysing and disseminating data on the magnitude of the OA impact advantage and the growth of OA across all disciplines worldwide (Brody, 2004).

The multiple online research impact metrics we are developing will allow the rich new database, *the Research Web*, to be navigated, analysed, 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? Is there work whose value is only discovered or rediscovered after a substantial period of disinterest? Can we identify the frequency and nature of such "slow burners"?

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 specialised 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 for Chapter 21 – to be confirmed and consolidated with the general list]

Adams, J. (2005) Early citation counts correlate with accumulated impact. *Scientometrics*, 63 (3): 567-581

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).

Alani, H., Dasmahapatra, S., O'Hara, K. and Shadbolt, N. (2003) Identifying Communities of Practice through Ontology Network Analysis. IEEE IS 18(2) pp. 18-25.

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

Berners-Lee, T., De Roure, D., Harnad, S. and Shadbolt, N. (2005) Journal publishing and author self-archiving: Peaceful Co-Existence and Fruitful Collaboration. http://eprints.ecs.soton.ac.uk/11160/

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,

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. (2006) Earlier Web Usage Statistics as Predictors of Later Citation Impact. *Journal of the American Association for Information Science and Technology* (JASIST, in press).

Connotea http://www.connotea.org/about

De Roure, D., Jennings, N. R. and Shadbolt, N. R. (2005) The Semantic Grid: Past, Present and Future. Procedings of the IEEE 93(3) pp. 669-681

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

Harnad, S. (1979) Creative disagreement. The Sciences 19: 18 - 20. http://www.ecs.soton.ac.uk/~harnad/Temp/Kata/creative.disagreement.html

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. and Brody, T. (2004) Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*, Vol. 10 No. 6

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., 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

Kleinberg, Jon, M. (1999) Hubs, Authorities, and Communities ACM Computing Surveys 31(4) http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/10.html

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

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

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

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

Newman , M. E. J. (2004) Coauthorship networks and patterns of scientific collaboration, *Proceedings of the National Academy of Sciences*. 101 suppl: 5200-5205

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

Okerson, Ann & O'Donnell, James (Eds.) Scholarly Journals at the Crossroads; A Subversive Proposal for Electronic Publishing. Washington, DC., Association of Research Libraries, June 1995

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

Skupin, A. (2004) The world of geography: Visualizing a knowledge domain with cartographic means. Proceedings of the National Academy of Sciences. 101 suppl. 1: 5274-5278

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

Sponsler, E. & Van de Velde E. F. (2001) Eprints.org Software: A Review. Sparc E-News, August-September 2001.

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

Wolfram, D. (2003). *Applied informetrics for information retrieval research*. Westport, CT: Libraries Unlimited.