ACCESS, USAGE AND CITATION METRICS: WHAT FUNCTION FOR DIGITAL LIBRARIES AND REPOSITORIES IN RESEARCH EVALUATION?

Chris Armbruster
Research Associate, Max Planck Digital Library,
Max Planck Society
Invalidenstrasse 35, D 10115 Berlin
www.mpdl.mpg.de

Executive Director, Research Network 1989
www.cee-socialscience.net/1989

Abstract

The growth and increasing complexity of global science poses a grand challenge to scientists: how to best organise the worldwide evaluation of research programmes and peers? For the 21st century we need not just information on science, but also meta-level scientific information that is delivered to the digital workbench of every researcher. Access, usage and citation metrics will be a major information service that researchers will need on an everyday basis to handle the complexity of science.

Scientometrics has been built on centralised commercial databases of high functionality but restricted scope, mainly providing information that may be used for research assessment. Enter digital libraries and repositories: can they collect reliable metadata at the source, ensure universal metric coverage and defray costs?

This systematic appraisal of the future role of digital libraries and repositories for metric research evaluation proceeds by investigating the practical inadequacies of current metric evaluation before defining the scope for libraries and repositories as new players. Subsequently, the notion of metrics as research information services is developed. Finally, the future relationship between a) libraries and repositories and b) metrics databases, commercial or non-commercial, is addressed.

Services reviewed include Leiden Ranking, Webometrics Ranking of World Universities, Ranking Web of Repositories, COUNTER, MESUR, Harzing Publish or Perish, Hirsch Index, CiteSeer, Citebase, SPIRES, SSRN CiteReader, RePEc LogEc and CitEc, Scopus, Web of Science and Google Scholar.

Keywords

Scientometrics, webometrics, research evaluation, research assessment, citation metrics, usage metrics, access metrics, digital library, digital repository, open access

Table of contents

I would like to thank Fredrik Astrom, Lars Björnshauge, Hakan Carlsson, Werner Marx, Emanuela Reale, Laurent Romary and Hermann Schier for extensive comment and criticism. For crucial input I also thank participants of a roundtable held on the subject of libraries, repositories and metric research evaluation at the University of Padua on 19 September 2007: Isidro Aguillo, Christoph Bruch, Johannes Fournier, Fred Friend, Wolfram Horstmann, John MacColl, Susanna Mornati, Sijbolt Noorda and Birgit Schmidt. A second roundtable was held at the University of Lund on 24 April 2008 with Frederik Astrom, Ahmed Hindawi, Wolfram Horstmann, John MacColl, Alain Monteil, Laurent Romary and John Wilbanks. Any remaining omissions and errors are mine.

1. Introduction: Advancing science by metric research evaluation

A century of unprecedented growth and internationalisation of research, higher education and scientific publishing has led to a degree of complexity that poses a grand challenge to scientists:

- How to keep track of very large amounts of information and tasks?
- What tools to use in organising knowledge and people in pursuit of new scientific breakthroughs?
- Which measures to take in enabling the further differentiation of research programmes while enhancing the global integration of researchers?

The Internet has principally eased access to knowledge while enhancing collaboration at a distance. Yet, the Internet also adds to the complexity, and the rise of e-science will reinforce this trend. Tools for self-observation are important and urgently needed.

Citation metrics have come to play an important role in many fields of the natural sciences, e.g. for hiring and tenure review as much as for decisions on to which journal to submit an article. Usage metrics are gaining ground in the social sciences. They are also relevant to research fields that have large numbers of users who are not also authors that will cite what they read. Moreover, the question whether earlier usage predicts later citations is being addressed. Citations and usage metrics are indicators for impact. Access metrics, by contrast, measure the online accessibility of research. This is not only a matter of volume, but also of quality. Access, usage and citation are distinct concepts, refer to different actions and are measured separately.

However, commercialisation has led to an impasse over access to knowledge (e.g., the anti-commons of patents, the serials crisis) and politically instigated research assessments and excellence awards lead to the rigid stratification of institutions. While research assessment typically has taken the form of peer review "by hand," government agencies have begun pushing for metric review as more efficient, impartial and objective. While political attention brings resources, it also carries the danger of arresting the further development of tools. Research assessment of institutions and individuals is one form of research evaluation, albeit a narrow one aimed at ranking and quality assurance.

Science has principally always been a world system, but evidently international databases could only be built through private funding, e.g., by the Institute for Scientific Information (ISI, now part of the Thomson Reuters Cooperation). However, the commercial definition of the corpus and the metrics that can be derived from it has historically been very narrow. Internal coverage of the corpus is sufficient only in those fields in which English language journal articles have for a long time been the overwhelming means of communicating ideas and research results. Practically speaking, metrics and related tools for tracking and indexing are presently most developed for physics, chemistry and biomedicine. Computer scientists and economists have been building not-for-profit online services for their communities, but these are not yet comprehensive and reliable enough.

Scientometricians have been reliant on commercial databases and tend to focus on performance measures. It may be surmised that this is due to a conjunction of zeitgeist (with targets and performance measures as a means of policy implementation and league tables for everything) and a receptive audience among research assessors (including government departments). However, the focus on past performance, league tables and top-up funding for alleged world-class research brings with it the risk of transforming the differentiation of research programmes into a rigid institutional stratification. The resulting social inequality that will turn ranking into a rat race might well undermine the essential cultural norms of science such as communalism, universalism, disinterestedness and organised scepticism (Merton 1942).

Current research evaluation is defined by political exigencies and constrained by commercial interests. However, so my argument, we need to understand that evaluation is something that researchers increasingly undertake on a daily basis. Moreover, due to the increasing global complexity of science we need new meta-level tools for self-observation. In this scenario, libraries and repositories will come to play an important role in developing tools and delivering data to the digital workbench of researchers.

Purpose and argument

My purpose is not to spuriously claim that digital libraries and repositories will deliver all the answers. The aim is principally to investigate the potential of new tools for research evaluation, focussing in the first instance on access, usage and citation metrics. Digital libraries have become involved in much more than the acquisition of scientific information, presenting and preserving materials produced by researchers, typically sharing them in open access (cf. California Digital Library for a university system or Max Planck Digital Library for a national research academy). At its most sophisticated, digital libraries bring materials to the researcher's workbench and capture and collect the products (e.g., data, articles). Digital libraries are increasingly significant at institutional, national, regional and global level (cf. the Digital Library Federation in the US, European Digital Library or World Digital Library).[1]

Digital libraries include repositories and digital repositories are a new type of library. Repositories come in various forms, having an institutional, national or disciplinary focus. Several large disciplinary repositories have been built for physics (arXiv), biomedicine (Pubmed), computer science (Citeseer), economics, law and the social sciences more generally (RePEc, SSRN). Interestingly, in biomedicine the disciplinary and national form is combined with the setting up of UK Pubmed. France has most consistently opted for a national database (Hyper Article en Ligne – HAL), while the US, Germany, UK, Japan and Australia together account for more than half the worlds' institutional repositories.[2]

Metric data may be aggregated to produce information about any unit associated with research – e.g., author, article, book, data set, funding, supervision, publication outlet, research team, institution and country. Provided that clean data

---

[1] http://www.cdlib.org/ - http://www.mpdl.mpg.de/ - http://www.diglib.org/ - http://www.edlproject.eu/ - http://www.worlddigitallibrary.org/

[2] http://arxiv.org/ - http://www.pubmedcentral.nih.gov/ - http://citeseer.ist.psu.edu/ - http://repec.org/ - http://www.ssrn.com/ - http://ukpmc.ac.uk/ - http://hal.archives-ouvertes.fr/ - http://www.opendoar.org/

are obtained according to a global standard, a federation of repositories would offer a database that would be superior in every respect: size, variety and quality. If we were to consider only citation metrics, a truly global database would emerge as digital libraries and repositories worldwide adhere to standardised descriptors for author, institution and item. Not just journal articles may be tracked, but also a variety of other publications, such as books, book chapters, conference proceedings, thesis, working papers, and possibly even patents. The version of any publication could also be controlled, distinguishing for example working papers from post-prints. This would produce a much richer impact on research as well as a much more nuanced picture of science.

A long and growing list of sceptical responses to metric research evaluation by individual researchers, journal editors and institutional leaders exists. As will become clear from the following analysis, current services do not respond yet to the grand challenge of enhancing the capacity of science for self-observation to such a degree that researchers can effectively reduce complexity in pursuit of the further advancement of knowledge. A popular line of argument is not to dismiss metric research evaluation outright, but to insist that experts (read: the senior scholars) must be the ultimate arbiters. However, on the grounds of logic and evidence there is little to suggest that individual experts or even a group of experts deliberating will reach a more informed understanding as to the importance or impact of research than the dispersed community that sets links, downloads and cites. This is not about the wisdom of the crowds, but based on the observation that an aggregate measure of access, usage and citation is a much more reliable predictor of influence than the opinion of experts.

That said, the following argument does not engage the critics of metric research evaluation. It is assumed that we need metrics for the further advancement of science. In that vein, we must search for the best possible solution. Currently, it is far from self-evident that the optimal solution will be implemented. Furthermore, we cannot even be confident that solutions adopted will further the advancement of science rather than obstruct it. Neither the existing commercial nor the non-commercial models hold enough promise and so-called open access scientometrics is also full of problems. With those reservations in mind, the following systematic exposition of digital libraries and repositories in relation to scientometrics addresses the grand challenge of furthering the future growth and advancement of science.

2. Libraries, repositories and scientometrics: enhancing the differentiation of research programmes and the global integration of scientists?

For metrics to be of utility for reading, writing and peer review, the sifting of applications, tenure review or the assessment of research groups, measures beyond stargazing are required. Because metrics lend themselves to rankings, there has been a marked tendency to construct top lists and focus all attention on the winners. While this is broadly compatible with the bestowal of recognition and reputation by peers, these kinds of rankings are one dimensional, about the past and often of research programmes long consolidated.

For ongoing research and teaching, measures need to be of a kind that signals contemporary relevance and future potential. This is important not only for further differentiation of research programmes, but also the global integration of scientists. One need only understand that competition and cooperation for grants and positions are increasingly global, and happen among scientists who have little or no prior face-to-face contact. Yet, as will become evident in this section, neither bibliometrics nor webometrics currently offers information that is fine-grained enough to be useful for scientists in their everyday practice.

Measuring impact: the bibliometrics of universities and departments

Scientometricians believe in the value of metrics for research evaluation (Moed 2005). As research funders increasingly demand research assessment, they receive more attention. As national research assessments are being based on metrics, for example in the UK and Australia, they attract more resources. National research assessments are the most comprehensive instance of research evaluation performed to date. It is thus important to understand what the present scope and quality of metric evaluation is.

The Leiden Centre for the Study of Science and Technology (CWTS) has developed a proprietary database derived from the Institute of Scientific Information (ISI), allowing for advancements in bibliometric methodology, such as defining a normalised citation impact for a number of scientific sub-fields. The normalised citation impact is measured by relating the average citation impact of a set of publications (by an author, team, department) to the world citation average in a given research field. Comparative research assessment of departments, institutes and teams is based on this measurement.

Leiden CWTS has investigated in how far the current ISI data could be the basis for reliable research evaluation, compiling a university ranking[3] and seeking to extend it to the departmental level. For the Leiden ranking much effort went into the careful definition, cleaning and unification of data, covering about 1/3 of worldwide journal publication, albeit “the most important international journals.” The Leiden ranking is restricted to universities and does not include, for example, national research academies. Among universities, the ranking is restricted to institutions with more than 5000 publications per annum, thus covering only the largest 400 universities with more than 600 researchers. That still produced one outlier, with a large number of publications but too little impact, which was excluded from the analysis.

The Leiden methodology provides a differentiated set of rankings, which are coded as follows (listed on the website)

- Yellow - ranking by size, i.e., number of publications (P)
- Green - ranking by the size-independent, field-normalised average impact (CPP/FCSm)
- Orange - ranking by the size-dependent ”brute force” impact indicator, the multiplication of P with the university’s field-normalised average impact

[3] http://www.cwts.nl/cwts/LeidenRankingWebSite.html

- Blue - ranking by the "simple" citations-per-publication indicator (CPP)

Yet, for the Leiden ranking to be valid and useful beyond a rough-and-ready public ranking of large science universities, research assessment would need to be possible at least at the departmental level. On behalf of the European Commission, Leiden CWTS and the Fraunhofer Institute for Systems and Innovation Research compiled departmental rankings for the life sciences (immunology, neuroscience, bioinformatics and genetics) and nanotechnology (Noyons 2003a, 2003b). However, a cost-effective and reliable way of counting, calculating and measuring excellence could not be devised. One important obstacle was that authors could not be identified unambiguously with their affiliation because address quality was too low.

Leiden CWTS was also charged with developing more reliable bibliometric indicators for computer science, based on the interjection of researchers that the Web of Science (WoS) is inadequate as a database because it does not cover (conference) proceedings, which in computer science are considered as important as journal articles (Moed/Visser 2007). An expansion of the WoS database through conference proceedings is feasible. However, even after the proceedings from several publishers were added, this boosted internal coverage only from 38% to 51%. For any research evaluation to be reliable, internal coverage should exceed 80% - which is the case only for subsets in biochemistry, molecular biology, human biology, clinical medicine, physics and astronomy. A comprehensive evaluation of science and medicine is only possible if one accepts internal coverage rates below 80%, in engineering below 60% and in the social sciences and humanities below 40%.

In expanding coverage beyond journal articles, they encountered the problem of version control. The problem exists on three levels:

a. Authors publish and re-publish the same item as technical report, conference paper and journal article, often with minor alterations only;
b. Authors seek to maximise publication count and potential impact by slicing their publications so as to report (portions of) an idea, key finding or methodology several times;
c. Only the publishers' version is considered authoritative but many contributions first circulate as work-in-progress.

Web of Science is a centralised database, reliant on the automatic extraction of well-formatted data. The extension of the database quickly runs into the problem that well-formatted (meta-)data are often not available. Additional costs therefore accrue to anyone wishing to use ISI data as results would need to be checked manually. This also means that a specialist is needed for interpreting results. Consequently, it is not surprising that much institutional and individual assessment is still done "by hand."

<u>Measuring access: webometrics of universities and repositories</u>

In the context of the rise of e-Science and the shift to electronic publishing, it is likely that metric research evaluation will be based on the Internet. A leading example is the webometric ranking compiled by the Consejo Superior de

Investigaciones Científicas (CSIC, Madrid). In constructing indicators, they adhere to the Berlin Principles on Ranking of Higher Education Institutions[4], which demand a clear purpose, transparent methodology and auditable data collection.

Principally, the Internet enhances the tracking of metrics for individuals and aggregates (i.e., journals, institutions, etc.) – totally and worldwide. To be sure, the Internet has a language bias and does not make people or institutions more equal, but everybody may be included. While Internet citation and usage metrics have become available (see below), reliable webometrics are available only at the aggregate level. Rankings are compiled for universities and repositories.

Webometrics are access metrics because they track not only impact, but also online volume and visibility of any domain as well as the availability of rich files. Webometrics are simultaneously about research excellence and public access. The Webometric Ranking of World Universities gives global and regional rankings, and includes a list of R&D centres. Institutions are ranked as such, using the domain name as referent. Validation has been achieved by correlation to standard university rankings. The Ranking Web of Repositories tracks the global visibility and impact of the repositories, listing the Top 300 that make available scientific publications and data.[5]

The indicators used in computing rank are defined as follows (listed on the website):

- Size (S). The page count is retrieved from Google, Yahoo, Live Search and Exalead. Results are log-normalised to 1 for the highest value of each engine. For each domain, maximum and minimum results are excluded and every domain is assigned a rank according to the combined sum.
- Visibility (V). The total number of unique external links received (in-links) is obtained from Yahoo Search, Live Search and Exalead. For each engine, results are log-normalised and then combined to generate rank.
- Rich Files (R). For institutions the following files are counted: Adobe Acrobat (*.pdf*), Adobe PostScript (*.ps*), Microsoft Word (*.doc*) and Microsoft Powerpoint (*.ppt*). These data are extracted using Google and the results for each file type are merged after log-normalisation. For repositories only text files (.pdf) extracted from Google and Yahoo are considered.
- Scholar (Sc). For institutions the number of items and citations for each academic domain is retrieved from Google Scholar. For repositories the mean of the normal number of papers and those published between 2001 and 2008.

For the institutions the four ranks are weighed to obtain a single overall ranking as follows:

Rank = 4*RankV+2*RankS+1*RankR+1*RankSc

For the repositories, the four ranks are weighed as follows:

Rank=5*RankV+2*RankS+1.5*RankR+1.5*RankSc

[4] http://www.cepes.ro/hed/meetings/berlin06/Berlin%20Principles.pdf

[5] http://www.webometrics.info/

Repositories and the rich files they contain positively affect institutional rank. If the Premier League is defined as Top 200, World Class as Top 500 and Regional Class as inclusion in the Top 1000 universities, then the thresholds in 2007 were as follows:

Table 1 – Rich file thresholds

	PDF	DOC	PPT	PS	SC (citations)
Premier League	19000	4000	2000	1000	3300
World Class	7000	2000	1000	300	1200
Regional Class	3000	500	300	50	400

Source: Communication from CSIC CINDOC

Webometrics are feasible. However, just as in bibliometrics, problems of address quality, version control and the crudeness of measurement exist. Webometrics is initially as much a process of evaluation as education. There are numerous bad practices in domain naming, content creation and conversion, interlinking and so on. Much content remains invisible. Older content vanishes. Much may be duplicate content. Moreover, webometrics provides at best a rough-and-ready ranking of institutions and repositories. As valuable as these are as guides, they are not sufficiently fine-grained for research evaluation.

<u>Opportunity and challenge for libraries and repositories: standardised descriptors for the online tracking of item, author and institution</u>

Science needs a metric evaluation system that starts not from the aggregate level, but the individual item, the author and her institutional affiliation. The groundwork for a universal metrics system is still missing: standard descriptors for item, author and institution. Worldwide tracking is easiest if a single standardised system of descriptors is used. It would be best if it was developed and owned by stakeholders that identify with a universal and long-term mission. Libraries and repositories come to mind. If they develop a universal standard, then commercial databases and service providers have a strong incentive to implement the standard.

By contrast, the development of metric research evaluation is likely to be hindered if competing (commercial) standards are established. Firstly, a commercial lock-in would mean that global tracking and seamlessly integrated services become subject to laborious agreements, which are unlikely, thus impeding the advancement of science. Secondly, science would become dependent on commercial entities and the future vagaries of mergers and bankruptcies. Thus it would be preferable if libraries and repositories were to develop a universal standard that is implemented and controlled locally.

The rise and prevalence of research assessment justifies the control and collection of metadata by institutional libraries and repositories, since assessment requires a complete record. Rather than relying on extraction only, institutional libraries and repositories will be in a position to record these automatically (e.g., in a publication management system) and control them manually. This will include a classification for type and version of the publication, e.g., working paper or book chapter, pre-print or published version. Local control distributes cost while enhancing quality.

A separate question is whether libraries and repositories will be in a position to generate access, usage and citation metrics from the content they hold. Webometrics is geared towards producing rankings (in keeping with current fashion) but also demonstrates that institutions benefit in terms of visibility if they hold content. This might not only be published content (or content intended for publishing) but also auxiliary material like conference slides or basic material like research data. Repositories should and will log data for the content they hold, only the creation of a reliable metrics database is not as straightforward an organisational task as collecting metadata on author, institution or type and version of publication.

3. Open science and research evaluation: repositories as base?

Research evaluation is often narrowly defined as research assessment, which rewards few and disappoints many. Moreover, metrics have been geared towards publishers' journals and are dependent on them. Metrics lend themselves to ranking. Rankings are popular because they relate information effectively: more/less; better/worse. In ranking performance, metric research assessment not only defines the rules of the game to which research organisations and individual researchers must adapt for the next round, but also frequently increases inequality.

Rankings are criticised because they are seen as entrenching existing inequalities of money and power. In the Ivy League, for example, the value of the endowment at Princeton, Harvard and Yale was more than $1 million per student by the year 2000, while at Brown, Dartmouth and Pennsylvania it was less than $200k (Ehrenberg/Smith 2003). As a cumulative effect, the top three universities may spend twice as much per FTE student annually ($80k) as the bottom three. This has significant consequences for faculty salaries and student bursaries, affecting recruitment and chances for initiating and promoting new research programmes, which compound to increasingly put the less rich at an ever greater disadvantage.

By contrast, open science is premised on widely distributing the means whereby scientific knowledge may be refuted or elaborated. However, as science has become a large and expensive system, the funders of science demand more efficiency. This has entailed a redistribution of scarce funds to "high performers." Open access has an important function in so far as it ensures that at least the basic means of participating in the scientific enterprise, i.e., access to research articles and data, is as universally available as possible, independent of institutional affiliation. What measures must be taken to ensure that metric research evaluation will not undermine but enhance the republic of science?

Repositories are viewed as highly compatible with open science. This is evident from the Open Access mandates that have been passed by research funders like the European Research Council (ERC), the National Institute of Health (NIH) or the UK Research Councils. Research institutions have also adopted mandates, notably the Faculty of Arts and Sciences of Harvard University. Open Access mandates are premised on the deposit of publications (and data) to repositories – disciplinary or institutional. What is their potential for research evaluation?

Institutional repositories for research evaluation?

Announcements have been made about the coming Open Access Scientometrics revolution (Shadbolt 2006, Harnad 2007). The vision has it that repositories and, in particular, institutional repositories are the “natural” place from which metrics are generated and displayed. It assumes a synergy between repositories, open access and metrics for research assessment, supported by research funders' open access mandates, which are believed to enhance impact for research articles and data. Research assessment is seen as a driver for the development of metrics more generally, as well as wider research evaluation services such as the mining and navigation of text and data.

Crucial to the vision is the assumption that the full corpus of research articles will become available through repositories. Indeed, the list of possible open access scientometrics services is impressive: service would ensure the disambiguation of names and addresses; count the number of downloads and citations; chart the growth rate, peak latency and longevity of impact; obtain scores for CiteRank, hub/authority, co-citation and co-text analysis; enable harvesting, inverting, indexing and citation-surfing; provide Boolean full-text search and ontologies; and even identify silent and unsung authors. Moreover, open access scientometrics is to deliver superior research assessment through a “dynamic multiple-regression engine, in which each of the individual vertical predictors can be given a weight and combined in a multivariate equation” – a perfectly calibrated battery of metrics (Harnad 2007).

This vision should and must be contrasted to the ongoing and not very successful efforts to persuade research assessment panels that repositories are a valid and reliable source. Software has been developed (Institutional Repositories for Research Assessment) to channel publications towards panels, but the UK RAE panels did not trust the veracity of the publications in the repositories and rejected their use. In response more software has been developed (Validating Repository Content) to certify that the author's post-print held in a repository is of the same value as the publisher's version, addressing the perceived “trust gap.”[6] The certificate is meant to attest the veracity of the fundamental facts and catalogues the differences between the version held by the publisher and by the repository. However, even if a certificate could be attached to every single article, the question remains - why would anyone trust and use these, if metrics derived from the formal publication are available?

[6] http://irra.eprints.org/ - http://valrec.eprints.org/

The idea that open access scientometrics may be built on the self-archived authors' version seems flawed. For reasons of accuracy, veracity and trust there is no scope for metrics that are not based on formal publication. Particularly research assessment related to funding, tenure and promotion (and its denial!) can justifiably only be based on originals. This would seem to leave institutional repositories with two possibilities, both of which may be pursued:

- Develop a strategy for the deposit of the published, authoritative version, for example, by working towards open access publishing;
- Focus on collaboratively developing standards for metadata that identify author, item and institution in a way that impact may be tracked worldwide.

<u>Metric services of disciplinary repositories</u>

Disciplinary repositories, in contrast to the struggling institutional ones, could deliver increasingly meaningful metrics because they have a well-defined corpus accumulated over many years. If we select the top three repositories according to the Ranking Web of World Repositories, what kinds of citation and usage metrics are available?

ArXiv (No. 1, Ranking Web of World Repositories)[7] is a fully automated e-print service and distribution system for pre-prints in physics and related disciplines such as mathematics, computing and quantitative biology. Founded in 1991, the service is currently hosted by Cornell University Library. In 2006 and 2007 an average of more than 4,000 submissions were processed each month. More than 480,000 papers have been deposited. However, neither citation nor usage metrics are available for the whole corpus. Some citation metrics for high-energy physics are provided by add-on services. SPIRES is a high-energy physics (HEP) database that indexes the relevant literature from arXiv, tracking citations (refers to, cited by) and providing overall top cited lists by year and cumulatively. For HEP the SPIRES services are of some value in tracking citations but not in counting as the service is not dynamic. Citebase is a semi-autonomous citation index that has tracked arXiv since 1999 but has access only to the UK mirror. Citation impact is not measured reliably.

Social Science Research Network (SSRN, No. 2)[8] features papers and abstracts in economics and law as well as related disciplines such as accounting, finance, negotiations, marketing, social insurance and management. Networks for political science and the humanities were set up in 2007. Accumulated since 1994, the SSRN repository features over 150,000 full-text documents from more than 90,000 authors. SSRN has counted more than 21 million downloads and is developing CiteReader (beta). The top author may boast more than 300,000 downloads, while No. 10 still has more than 80,000 downloads and No. 50 has more than 32,000 downloads. The top paper has more than 64,000 downloads, No. 10 more than 22,000 and No. 50 more than 7,900. SSRN usage metrics are reliable and sophisticated. Reliability relates not only to the discounting of robots and automated access, but also to the distinction between abstract views and actual downloads - users need to click on the abstract to be able to download. SSRN

---

[7] http://arxiv.org/ - http://www.slac.stanford.edu/spires/hep/ - http://www.citebase.org/
[8] http://www.ssrn.com

displays usage statistics for every paper and provides current and cumulative top lists for a multitude of subject areas as well as, on aggregate, for research networks. CiteReader promises not only citation tracking but also a measure of impact. References and footnotes of SSRN papers have been resolved and more than 77,000 papers have cites. However, with the CiteReader SSRN runs into the problem of needing a well-defined corpus for the citation count to be meaningful. SSRN has strong areas like economics, business and law but has been designed to cover the social sciences broadly, now also including the humanities. While broad coverage attracts users, it poses numerous problems for citation metrics: internal coverage is too low, field normalised averages would need to be established and version control imposed. While the usage metrics are meaningful, any measurement of citation impact is not reliable.

Research Papers in Economics (RePEc, No. 3)[9] links over 750 institutional repositories, holding 237,000 working papers and over 344,000 journal articles. 478,000 items are available online. RePEc services (LogEc and CitEc) count abstract views, downloads and citations for working papers, journal articles, books, book chapters and software items. Reliability is enhanced by removing access by robots and spiders as well as implementing measures to avoid double counting. As with SSRN, usage statistics indicate the relation of abstract views to downloads. Citations RePEc has been tracking for some years and internal coverage has been assessed. Utilising a list of the top 1000 economists, RePEc covers 72% of the authors and 61% of citations for the period 1990-2000. While the top 1000 economists are taken from across the full spectrum of economics and management, RePEc concentrates on economics and finance. Also, RePEc only started in 1993. Because RePEc has a well-defined and limited corpus, one would anticipate that for the decade from 2000 onwards, internal coverage exceeds 80%. For economics and finance a field-normalised average can be established. Some version control (working paper versus journal article) is given. RePEc demonstrates that repositories (disciplinary, federated) are principally a viable alternative to collect usage and citation metrics.

Opportunity and challenge for digital libraries and repositories: setting standards for statistics

When looking at repositories, one needs to acknowledge the distinction between content, metadata and metric data. Repositories are built to capture content. By contrast, commercial citation databases are not based on captured content, but on the extraction of metadata. In the section above it has been suggested that the first challenge is for libraries and repositories to define a universal set of descriptors for the metadata. The second challenge then is to come up with a shared standard for the collection of statistics.

Repository managers need to understand that the self-archiving of content (so-called Green Open Access) does not enhance the value of repositories for metric research evaluation but heightens the problem of version control. If content is sought to build up a corpus, then the better strategy is to focus on working paper series (pre-prints) and the published, authoritative version. Metrics for working

[9] http://repec.rog

papers are interesting, especially usage statistics, because they give authors early feedback, journal editors additional information and readers an easy way to navigate the new and emerging literature.

Yet content and metrics may also be collected and held separately. Libraries and repositories are the only place that may aspire towards holding metadata on the full scientific record, including even artistic expressions in the humanities. If these metadata are expressed in a universal standard, then metric evaluation may be added on as a research information service. For these research information services, in turn, standards need to be defined so that the statistics are reliable.

4. Metric evaluation: towards research information services

For metrics to be of value to scientists, they must be available on a daily basis. There is substantial anecdotal evidence that

- Review panels for major grants, often consisting of the most respected scientists in a field, routinely use existing metric information services to gather information about applicants;
- Hiring and tenure committees routinely evaluate candidates and faculty against citation metrics;
- Government and university induced assessments of departments and research groups are increasingly based on metrics of some kind.

Principled objections regarding the validity and reliability of research metrics exist (Adler 2008). A system of identifiers and standards for statistics, as suggested above, would significantly increase reliability. Ongoing scientometric research must address the validity of indicators. The discussion below seeks to outline ways of achieving more sophisticated indicators and promote the informed use of metrics.

<u>Citation metrics and the example of the h-index as research information service</u>

The h-index, proposed by J. Hirsch, has been very well received due to its convergent validity with existing bibliometric indicators as well as peer review, its robustness as a measure of lifetime achievements and the simplicity of its calculation in ordering papers by "times cited." The h-index is defined as follows (Hirsch 2005):

> *"A scientist has index h if h of his/her N_p papers have at least h citations each, and the other (N_p-h) papers have no more than h citations each."*

Indices are usually discussed for their advantages and disadvantages, and the h-index has limitations such as being only retrospective and not predictive and comparability of individual h-indices is given only for a limited and well-identified set of authors. However, what is remarkable about the h-index is that discussion has led to a set of modifications, transformations and extensions of the h-index so that a set of indicators has become available that, from a variety of angles, enables more informed decisions about the direction of one's own research.

Modifications of the h-index are primarily geared to shift the focus from lifetime achievement towards contemporary relevance and future potential. At the individual level these modified h-indexes may be collected for published authors to gauge originality, contemporary relevance and the capacity to set a trend. Though these modifications cannot do away with the Matthew effect in the citing behaviour of the peers, they do help to level the playing field between senior and junior authors by weighing and truncating publications and citations in a manner that discounts mere output due to age in the field. At the aggregate level of research groups and departments, these modifications would have the same effect, being fairer to younger research groups or new and upcoming departments.

Table 2 - Modifications of the h-index

	Contemporary h-index	Trend h-index	Creative h-index
Proposed modification	h-index weighted by age of the article, the coefficient γ set at 4, i.e., an article uses value with age, e.g., from 4/1 for citations in the first year to 4/10 in the tenth year.	h-index weighted by the age of the citation, the coefficient γ set at 4, i.e., a citation uses value with age, e.g., from 4/1 for citations in the current year to 4/10 ten years ago.	Truncation of publication network to n references made and m citations received.
Significance of modification	Moves focus from lifetime achievement to cotemporary relevance.	Discloses lasting influence as well as new trendsetters.	Discloses originality as a function of n and m, weighted for the number of authors per item, but principally cumulative for the individual or aggregate observed.
Source citation	A. Sidiropoulos, D. Katsaros and Y. Manopoulos 2006	A. Sidiropoulos, D. Katsaros and Y. Manopoulos 2006	J.M. Soler 2006

Source: as cited above

Transformations of the h-index are proposed with the intention to shift the focus of measurement. The common feature of these transformation is that they shift the focus from lifetime achievement to performance in a given number of years, seeking to identify world class and, where possible, stellar performances.

Table 3 – Transformations of the h-index

	g-index	R- and AR-index	b-index
Proposed transformation	Transformed by giving more weight to cumulative and absolute number of citations, i.e., g-index is the largest number such that the top g articles together receive at least g^2 citations.	All publications between rank 1 and h (the Hirsch core) are appraised for citation intensity (R-index) or weighted by age (AR-index, using a root function).	Transformed by introducing a baseline and using percentile rank scores as a standard for comparison in a given field.
Significance of transformation	The g-index measures global citation performance and introduces steeper gradient among highly-cited individuals or aggregates.	The R-index gives due recognition to very highly cited papers, highlighting the exception al contribution of a specific item, whereas the AR-index highlights highly cited papers that are new.	The b-index gives the number of papers in the set of an individual or aggregate that fall into the top x% (e.g., 10% for world class) in any field, based on the number of citations over n years (e.g., assessment cycle).
Source citation	L. Egghe 2006	B. Jin, L. Liang, R. Rousseau and L. Egghe 2007	L. Bornmann, R. Mutz and H.-D. Daniel 2007

Source: as cited above

Proposals have also been made to extend the h-index to journals, conference proceedings, topics and objects. Despite the known limitations of the Journal Citation Reports and the Journal Impact Factor, it is unclear whether the extension of the h-index to journals and conferences offers any substantial advantages. However, the extension of the h-index to topics (e.g., search strings) and objects (e.g., chemical compounds) is promising.

Table 4 - Extension of the h-index to topics and objects

	Topics h-b index	Objects h-b index
Proposed extension	To topics, hb ~ mn (m is the gradient of citations in n years since research began).	To objects, hb ~ mn (m is the gradient of citations in n years since research began).
Significance of extension	Differentiate new from older and possibly spent research programmes, e.g., by using search strings.	Differentiate new and under-researched objects from well- and possibly over-researched ones - e.g., chemical compounds, digital objects.
Source citation	M.G. Banks 2006	M.G. Banks 2006

Source: as cited above

While the h-index was proposed as a performance measure, the above set of modifications and extensions have turned it from a single measure into a plurality of indicators. As a research information service, the h-index becomes less vulnerable to the unsolved riddle of what citation counts actually measure

(Bornmann/Daniel 2007). In the normative view of science, one that performance measurement necessarily endorses, citations are an expression of recognition for the advancement of knowledge. Trivialised, this becomes the notion that awarding funding on the basis of past citations secures future advancements of knowledge. Constructivist investigations of science have shown that the motivations for citing are many, including the desire to establish connection to peers, persuading readers with references to authorities, flatter journal editors and likely peer reviewers and promote one's own work. Performance measurement corrects for self-citation but must ignore other motivations. This insistence on the normative view of science and the assumption of "disinterested" citation behaviour is very remarkable given how competitive publishing, funding and tenure have become and how much credence the view of homo sapiens as utility-maximiser has achieved.

Insofar as the h-index is understood as a research information service, the citations that it counts are a measure of whom and what colleagues pay attention to and where research has been cumulative (e.g., topics and objects). Indeed, for anybody embarking on a new line of research, be it as doctoral student, post-doctoral researcher or principal investigator, the various h-indices measures provide valuable information about topics, people and trends.

Usage metrics and the example of library statistics

View and download figures have been provided by repositories to authors and readers, demonstrating the value of their services. Counts are provided for how many times an abstract has been displayed, a page viewed and a document downloaded. It has also been essential that browsing is distinguished from actions that are taken as a proxy for usage, e.g., the online reading or downloading of the full research article. The validity of this distinction is ensured if online reading or downloading requires of the user a conscious request for the research article after having seen summary information such as title, author, abstract and so on. Larger repositories have provided robust figures since the 1990s. Indeed, service providers ensure that data is as reliable as possible by disabling and discounting robots, search engines, repetitive downloads and so forth (cf. Mayr 2006).

Usage metrics have become widespread, much like bestseller lists. In contrast to citation metrics, which require extraction and are dependent on later publications, online usage metrics are automatically recorded in the present for any item that is accessed. Usage metrics generated by readers are highly inclusive, easily accessible and inexpensive to collect and aggregate (Bollen/van den Sompel 2006, 2008). Usage metrics have potential as research information service. They are seen as complimentary to citations metrics. Some believe that earlier usage can predict later citations (Brody/Harnad/Carr 2006). In the relationship between publishers and libraries, usage metrics are deployed to evaluate subscriptions and price. For research performance, usage metrics are particularly relevant for fields in which knowledge is much applied and used, but not so much cited. For junior authors, usage metrics will be available first. For all authors usage metrics are available earlier if work-in-progress (e.g., working papers, pre-prints) is posted to repositories.

The potential of usage metrics resides in obtaining more information about readers' behaviour, judgement and uses. Digital libraries, as points of access and dissemination, are prime sites for the generation of usage metrics. Standardisation of usage metrics has been pushed by a coalition of publishers and libraries. Usage metrics from large disciplinary repositories have value, while statistics from institutional repositories will be meaningless unless they are aggregated in community services: the corpus has to be distinct, relevant and large enough (e.g., arXiv, SSRN and RePEc).

Basic standardisation has been achieved since 2002 by the project Counting Online Usage of Networked Electronic Resources (COUNTER).[10] COUNTER is a membership organisation of libraries and vendors. Libraries observe usage patterns and measure the value of service (i.e., subscriptions and prices in relation to usage), while publishers observe usage patterns and employ a standardised metric for product development and sales. For journals and databases, metrics are reported for

- Successful full-text article requests (differentiated for HTML and PDF)
- Turnaways
- Total searches and sessions

Vendors generate the metrics. The ambition is to establish a journal usage factor that complements the impact factor.

For digital libraries prototyping is undertaken by the project Metrics from Scholarly Usage of Resources (MESUR)[11], begun in 2006 at the Los Alamos National Laboratory. MESUR aims to provide a formal model of scholarly communication and representative usage data, which are linked to citation data and the COUNTER usage metrics for the purpose of cross-validation. The emphasis is on the large-scale definition and implementation of usage metrics. Data are obtained from a variety of universities, aggregators and publishers. Results from MESUR may be expected for late 2008.

Usage metrics have gained credibility. However, none of the disciplinary repositories or open access publishers is a member of COUNTER, which indicates a struggle over the definition of usage metrics, with the publishers clearly being interested in defining usage metrics narrowly, which in practice would limit metrics to that which publishers provide online: journals and their articles. A journal usage factor may lock in usage metrics with citation metrics as something that is provided commercially and controlled by publishers. MESUR, by contrast, points to a different future for usage metrics: data is generated not only at the source (e.g., publication) but also at the point of entry or access (e.g., digital library, search engine and so on). Moreover, the COUNTER metrics remain the most simple and limited, while MESUR points to a large variety of metrics that serve to map the structure of science and trace the differentiation of research programmes.

---

[10] http://www.projectcounter.org/
[11] http://www.mesur.org

<u>Metric batteries as research information service</u>

Metric batteries have been developed by repositories. They result from the proliferation of indicators. As the following examples show, construction of metric batteries is simple enough and the methodology is clear. The unresolved issues relate to the corpus on which the metrics are based, whether persistent identifiers are available for author, item and institution and what the standards are for the collection of the statistics.

The RePEc metrics battery is at least based on a reasonably coherent corpus whose internal coverage is likely to exceed 80% and effort has gone into defining statistics collection standards (as shown above). Besides RePEc, some of the other repositories discussed here, like arXiv, Citeseer and SSRN potentially have within them a corpus (e.g., high-energy physics, economics and management) that would lend itself to building research information services. However, the quantity and quality of content and metadata would need scrutinising.

Box 1 – The RePEc battery of citation, journal and usage metrics

Citations metrics, weighted for the prestige of the citation as well as multiple authors	• Number of Citations • Number of Citations, Weighted by Simple Impact Factor • Number of Citations, Weighted by Recursive Impact Factor • Number of Citations, Weighted by Number of Authors • Number of Citations, Weighted by Number of Authors and Simple Impact Factors • Number of Citations, Weighted by Number of Authors and Recursive Impact Factors • h, where author has written h papers that have each been cited at least h times
Journal metrics, based on number of pages and weighted for prestige of citation and multiple authors	• Number of Journal Pages • Number of Journal Pages, Weighted by Simple Impact Factor • Number of Journal Pages, Weighted by Recursive Impact Factor • Number of Journal Pages, Weighted by Number of Authors • Number of Journal Pages, Weighted by Number of Authors and Simple Impact Factors • Number of Journal Pages, Weighted by Number of Authors and Recursive Impact Factors
Usage metrics, distinguishing between abstract views and downloads, weighted for multiple authors	• Number of Abstract Views in RePEc Services over the past 12 months • Number of Downloads through RePEc Services over the past 12 months • Number of Abstract Views in RePEc Services over the past 12 months, Weighted by Number of Authors • Number of Downloads through RePEc Services over the past 12 months, Weighted by Number of Authors

Source: RePEc website

Google Scholar is a database that can be used by anyone for the metric evaluation of science. Indeed, anyone can now evaluate anybody with the help of a piece of software, "Publish or Perish," that is freely available on the Internet. There is anecdotal evidence of its widespread use, for example, by candidates who did not get hired in making the case that the candidate who did get hired did not have a track record of "impact." The idea that candidates may use metrics to challenge the decisions of hiring or tenure committees will make many uneasy. It does demonstrate, however, that metric research evaluation is taken up intentionally and is here to stay.

Box 2 – Harzing's software for citation metrics

Publish or Perish

Publish or Perish is a software program that retrieves and analyzes academic citations. It uses Google Scholar to obtain the raw citations, then analyzes these and presents the following statistics:

- Total number of papers
- Total number of citations
- Average number of citations per paper
- Average number of citations per author
- Average number of papers per author
- Average number of citations per year
- Hirsch's h-index and related parameters
- Egghe's g-index
- The contemporary h-index
- The age-weighted citation rate
- Two variations of individual h-indices
- An analysis of the number of authors per paper

Page link: http://www.harzing.com/pop.htm
Version: 2.4.2894 (3 December 2007)

Source: Harzing's website as cited above

Opportunities and challenges for digital libraries and repositories: building research information services

Metric batteries and anecdotal evidence about their use by scientists at the highest level indicate the principal value of access, usage and citation metrics. However, metrics would need to be implemented as a tool not only to assess institutions and researchers but also, and more importantly, to serve the intellectual differentiation and social integration of research programmes on a global scale. The wider significance of research evaluation becomes clear when it is understood that metrics are just one tool in research evaluation, alongside other tools such as data and text mining or concept formation and semantic enrichment. Metric tools would not only span access, usage and citation, but also include tracking and indexing. A plurality of metrics enables researchers to observe others and themselves from a variety of angles. Metric plurality does not open the floodgates of manipulation. To the contrary, a democratic polity and open science are not compatible with a single measure that reduces research to performance only. Moreover, a plurality of metrics enables researchers to state autonomously and individually their case to colleagues and funders.

The example of the h-index shows the potential of and demand for research information services that are usable on a day-to-day basis. It is hard to exaggerate the significance of this. Detractors may point to potential abuse by evaluators, to gaming strategies by authors and the prevalent unreliability of citation metrics with respect to coverage and author identification, but researchers will not desist from using the h-index and its variants. Indeed, the premise of the present argument is that these kinds of research information services have become indispensable. Criticising abuse and misuse of statistics is important, but the challenge is to build better research information services.

Libraries and repositories already play an important role in collecting usage and citation metrics. This works both ways as the library or repository may collect statistics for materials accessed through the library as well as from the library and repository. What needs further strategic reflection is what kinds of services might be offered to researchers and how libraries and repositories position themselves vis-à-vis existing databases and search engines that offer some metric indicators already.

5. Conclusion: What is the relationship between digital libraries, repositories and metrics databases?

Metrics and related services like indexing and tracking require a database. While a global usage database does not yet exist, for citations there are three highly functional ones: Web of Sciences, Scopus and Google Scholar. If libraries and repositories are to become players, then it is essential to understand the scope for competition and collaboration.

High functionality, limited value and the question of compatibility: commercial databases and search engines

Citation metrics require a large corpus. Elsevier's launch of SCOPUS in 2004 presented an alternative to Web of Science. Google Scholar also offers a database but the extraction of citations is not controlled by the definition of a particular corpus. Note that the corpus of SCOPUS is also not particularly well defined. Limited functionality is provided by EBSCO ASE and Highwire Press as multidisciplinary databases and Citeseer, PsycInfo, Pubmed Central, RePEc Econpapers, SciFinder Scholar and SMEALsearch as subject-based databases (cf. Bosman, et al. 2006). None of these services would seem to be in a position to rival WoS, Scopus and Google for global citation metrics, but they may have a future in their niche for related services such as citation indexing and may also come to play an important role in usage metrics as these databases typically also hold content.

WoS and Scopus are principally open to including open access publications and, in particular, open access journals, but they have no reason to cooperate with repositories. Google Scholar and other search engines include content from repositories, which are directly accessible. The compatibility of repositories with search engines is high but there is no compatibility with the large commercial databases. A comparative evaluation may help in determining if and how repositories could be valuable for metrics delivered through search engines and databases. Comparative studies have been undertaken already, but with different questions in mind (Jasco 2005, Bakkalbassi, et al. 2006, Bosman, et al. 2006). Nevertheless, it is possible to reuse the results to address questions of functionality and value and thereby to gauge the chance that repositories and libraries would become a significant source of metrics. As global databases exist for citations, the comparison will be limited to this field – but the implication for usage metrics is analogous. The categories of comparison for citation databases may be defined as follows:

- Functionality is defined by the availability of (meta-)data, the interface for searching, linking and tracking and the speed with which results are produced.
- Coverage refers to the absolute number of documents that are included directly or indirectly (e.g., via citation listing or tracking), the relative coverage in relation to subject areas and document type, the period covered and the speed with which the database is updated.
- Quality is determined by the nature of (meta-)data included per document (e.g., author, affiliation, title, abstract, keywords, publication type, citations, etc.) and their cleanliness and reliability.
- Ease of use is largely determined by the functionality, coverage and quality but in good measure also related to the experience a user has.

Repositories have no relationship with commercial databases but are harvested by search engines. For the users of search engines, repositories are an open access boon. Can repositories forge a relationship with one or more search engines that will lead to better and more accurate displays of usage and citations? Which search engine is most willing to invest in improving the quality of hits that scientists value so much in WoS and Scopus?

Table 5 - Comparison of functionality, coverage, quality and ease-of-use for large citation databases

Functionality	- WoS and Scopus have a corpus. WoS even has a clear rationale for defining the corpus, highly cited journals, whereas the Google Scholar corpus is not well-defined but more expansive. - Scopus is more versatile than WoS, particularly when it comes to author identification and citation tracking, but Google Scholar equally provides identification and tracking. - Google Scholar significantly outperforms WoS and Scopus not only in ease of access but also in the display of results.
Coverage	- WoS and Scopus are unrivalled when it comes to coverage of international journal articles, but for historical and commercial reasons their coverage is skewed towards STM journal articles, whereas Google Scholar principally covers all of the sciences and humanities and every type of item. - In a direct comparison of WoS (9,000 journals) and Scopus (15,000 journals), the coverage of Scopus is 5-15% smaller across subjects before 1996, but 20-45% larger after 1996.Ccoverage in Google Scholar is much broader, also internationally. - Google Scholar is significantly better at keeping the database current, as the newest literature is often instantaneously available, whereas WoS and Scopus experience a time lag.
Quality	- WoS and Scopus have significant added value for professional use. Because of the quantity of information available for every document (e.g., DOI, copyright, citations) their quality is unrivalled except for one major gap: research data. Niche providers (e.g., chemical compounds, geospatial data) fill this gap. - Keyword searches are handled slightly differently with Scopus offering a controlled vocabulary (thesaurus) and WoS generating keywords from references. - With regard to citation tracking, experimental searches have shown an overlap of 80-90% between WoS and Scopus, but a comparison with Google Scholar shows significant divergence with references overlap being as low as 1/4 and often not significantly better than 50%. The major limitation of WoS and Scopus with regard to the quality of searches is their limited scope when it comes to languages and to different document types.
Ease-of-use	- Google Scholar significantly outperforms WoS and Scopus in terms of speed, needing only 1-2 seconds to build a page in response to a search query, whereas the others need between 8 and 13 seconds. - In terms of the quality of hits, natural scientists are wholly satisfied with Scopus but largely dissatisfied with Google Scholar. The situation is different for the social sciences and humanities, but the general picture is that for those research fields in which the commercial databases hold content, Scopus is slightly better than WoS and both significantly more accurate than Google Scholar. - The interface of Scopus is lauded for its structure, ease of navigation and options for search refinement. WoS is perceived as requiring more roundabout actions to obtain results. Unique to Scopus is a citation overview that provides citations by article for each author with a good degree of reliability.

Source: Compiled from Jasco 2005, Bakkalbassi et al. 2006, Bosman et al. 2006

Stifled potential: non-commercial databases

Non-commercial databases invariably offer search functions and some have citation indexing and tracking. It is often assumed that what stands in the way of enhanced functionality and quality is the lack of journal articles available in open access. However, a critical experiment has shown that these databases already have problems with coverage even if items are available in open access. It has been found (Bergstrom/Lavaty 2007) that for 33 key economic journals, ninety percent of articles in the most-cited journals had been self-archived and about fifty percent of articles in less-cited journals were also available freely online. All of the freely available articles were found through Google. Using Google Scholar, about 10% less was found. However, when using OAIster only 1/4 of the freely available articles were retrieved and results were only marginally better for SSRN and RePEc searches. Given the high propensity of economists to self-archive and the availability of institutional and disciplinary repositories, the differences between Google and the non-commercial solutions are so dramatic as to warrant the conclusion that the non-commercial solutions, whatever their merits, will have only very limited potential in the future.

Further evidence to support the conclusion of the limited potential of non-commercial solutions is available:

1. RePEc (with LogEc and CitEc) offers a search facility to track titles and find a freely available version, but apparently this was to no avail for Bergstrom and Lavaty. RePEc has built-in services that log usage data and track citations. An impressive battery of metrics is available. Yet all services are experimental only and there seems to be no funding available for their further development. Even if funding were to become available, there is no business model in place that would ensure that services would continue to be developed. The reason that RePEc remains functional and viable is that a large number of economics departments and other institutions post their working paper series to RePEc, meaning that new findings become available well before they are published in a journal.
2. Citebase is a citation tracker for arXiv. Closer examination reveals that while funds were once available to build Citebase, it has access only to the UK mirror of arXiv, not to the main host and not to any other mirror. In effect, Citebase languishes and it is unclear if it will ever achieve full functionality for arXiv, let alone any other research field.
3. Citeseer did achieve functionality for the research field that it covers because it was implemented as the core service (Autonomous Citation Index). However, new commercial rivals have emerged and prospered (Google, Scopus) and additional functionality has been developed for publisher platforms. Citeseer was unable to match these investments initially but has recently attracted more funding, so that Citeseer continues to enjoy comparable functionality. As WoS focuses on the published journal literature only and in computer science a significant portion of publishing is done through conference proceedings, Citeseer covers an exclusive niche.

Strategic alternatives for libraries and repositories: club good or search engine?

The significance of libraries and repositories (of institutions, countries, disciplines and publishers) is that they provide the opportunity for fuller coverage and better metadata quality (e.g., non-ambiguous identification of authors with their institutions). However, repositories cannot do this individually. Service provision requires collaboration with an established database or the building of a new one. Non-commercial databases have conducted important experiments and demonstrated potential, but alone they lack the capacity for further investment. However, libraries could opt to fund present-day services, such as CiteSeer and RePEc, and then seek to expand the model to other research fields.

More generally, how could metric research information services be organised? Two alternative solutions come to mind:

A. The provision of metadata and metric data to commercial search engines (e.g., Google Scholar, Scirus) on a nonexclusive basis in a public-private partnership that enhances the functionality of search engines. The rationale would be that exposure enhances impact and therefore increases the return on investment on research, which justifies the public expense. Access would be immediate and free through the search engine.
B. The creation of a club good by research funders, organisations and libraries, i.e., a database as an overlay service, for the use of which those institutions that provide clean and reliable metadata are rewarded with access at “production cost” whereas all other users are charged a higher usage fee (to eliminate the free rider problem). This commercially self-sustaining (and not-for-profit) service would offer clear incentives for institutions and repositories around the world to join and this would expand access universally.

Solution B would be feasible if institutions were to join forces and membership is limited to institutions that meet certain quality criteria. By creating an overlay service as club good, this good has the initial exclusivity that makes it attractive to join the club. The exclusivity is not a means to exclude others, but a way of ensuring working standards on the basis of which club membership may be extended universally. The critical factor in making B work is that enough prestigious institutions commit initially. In the absence of B, any individual institution and repository would be better off opting for A, so that it may better demonstrate its impact to funders and colleagues. The largest benefit will accrue to the early movers and those that can utilise their back file.

In sum, libraries and repositories face a challenge, which provides them the strategic opportunity to

- Develop a system of descriptors that will persistently and unambiguously identify item, author and institution;
- Develop universal standards for the collection of access, usage and citation statistics;
- Examine the feasibility of building research information services to advance science to the next frontier.

Cited literature:

Adler, R.; Ewing. J., Taylor, P. (2008) Citation Statistics. A joint report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). Joint Committee on Quantitative Assessment of Research.
http://www.mathunion.org/publications/report/citationstatistics/

Bakkalbassi, N. et al. (2006) Three options for citation tracking: Google Scholar, Scopus and Web of Science. Biomedical Digital Libraries 3 (7)

Banks, M.G. (2006) An extension of the Hirsch index: Indexing scientific topics and compounds. Scientometrics 69 (1) 161-168

Bergstrom, T.C., Lavaty, R. (2007) How often do economists self-archive? *Department of Economics, UCSB.* http://repositories.cdlib.org/ucsbecon/bergstrom/2007a

Bollen, J., Van den Sompel, H. (2006) An architecture for the aggregation and analysis of scholarly usage data. ACM: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries, 298-307

- (2008) Usage Impact Factor: the effects of sample characteristics on usage-based metrics. JASIST 59(1)1-14

Bornmann, L., Daniel, H.-D. (2007) What do citation counts measure? A review of studies on citing behaviour. Journal of Documentation, forthcoming

Bornmann, L., Mutz, R., Daniel, H.-D. (2007) The b index as a measure of scientific excellence. A promising supplement to the h-index. Cybermetrics 11(1) 6

Bosman, J. et al. (2006) Scopus reviewed and compared. The coverage and functionality of citation database Scopus, including comparisons with Web of Science and Google Scholar. Utrecht University Library

Brody, T., Harnad, S., Carr, L. (2006) Earlier Web usage statistics as predictors of later citation impact. Journal of the American Society for Information Science and Technology 57(8) 1060-1072

Egghe, L. (2006) Theory and practise of the g-index. Scientometrics 69 (1) 131-152

Ehrenberg, R.G. and C.L. Smith (2003) What a difference a decade makes: Growing wealth inequality among Ivy League institutions. Cornell: CHERI working paper

Giles, C.L., K.D. Bollacker and S. Lawrence (1998) CiteSeer: An Automatic Citation Indexing System - Giles, Bollacker, Lawrence (1998). Digital Libraries 98 - Third ACM Conference on Digital Libraries, Edited by I. Witten, R. Akscyn, F. Shipman III, ACM Press, New York, pp. 89-98,

Harnad, S. (2007) Open Access Scientometrics and the UK Research Assessment Exercise. In: Torres-Salinas, D. and Moed, H. F., (eds.) Proceedings of 11th Annual Meeting of the International Society for Scientometrics and Informetrics 11(1), pp. 27-33, Madrid, Spain.

Hirsch, J. (2005) An index to quantify an individual's scientific research output. PNAS 102(46) 16569-16572

Houghton, J.W. and P.J. Sheehan (2006) Economic Impacts of Enhanced Access to Research Findings. CSES Working Paper
http://www.cfses.com/documents/wp23.pdf

Jasco, P. (2005) As we may search – comparison of major features of the Web of Science, Scopus and Google Scholar citation-based and citation-enhanced databases. Current Science 89, pp. 1537-1547

Jin, B., Liang, L., Rousseau, R., Egghe, L. (2007) The R- and AR-indices: Complementing the h-index. Chinese Science Bulletin 52(6) 855-863

Mayr, Philipp (2006) Constructing experimental indicators for open access documents. Research Evaluation 15(2)127-132

Merton, R.K. (1942) The Ethos of Science. Republished in On Social Structure and Science. Edited by P. Sztompka (1986) Chicago: University of Chicago Press

Moed, H.F. (2005) Citation Analysis in Research Evaluation. Dordrecht: Springer

Moed, H.F., Visser, M.S. (2007) Developing Bibliometric Indicators of Research Performance in Computer Science: An Exploratory Study. Research Report to the Council for Physical Sciences of the Netherlands Organiation for Scientific Research (NOW). Leiden CWTS

Noyons, E.C.M. et al. (2003a) Mapping Excellence in Science and Technology across Europe. Life Sciences. Leiden CWTS and Fraunhofer ISI

Noyons, E.C.M. et al. (2003b) Mapping Excellence in Science and Technology across Europe. Nanoscience and Nanotechnology. Leiden CWTS and Fraunhofer ISI

Sidiropoulos, A., Katsaros, D., Manolopoulos, Y. (2006) Generalized Hirsch h-index for Disclosing Latent Facts in Citation Networks. Scientometrics 72(2) 253-280

Shadbolt, N., Brody, T., Carr, L. and Harnad, S. (2006) The Open Research Web: A Preview of the Optimal and the Inevitable. In: Jacobs, N., (ed.) Open Access: Key Strategic, Technical and Economic Aspects, chapter 20. Chandos.
http://eprints.ecs.soton.ac.uk/12453/

Soler, J.M. (2006) A Rational Indicator of Scientific Creativity. Journal of Informetrics 1(2) 123-130