



Law, Derek (2007) *Making science count: Open Access and its impact on the visibility of science*. In: Institutional Archives for research. Experiences and projects in open Access. Proceedings. Rome Nov. 30 - Dec 1, 2006. Rapporti ISTISAN. Istituto Superiore di Sanità, Rome.

<http://eprints.cdlr.strath.ac.uk/2841/>

This is an author-produced version of a paper published in Institutional Archives for research. Experiences and projects in open Access. Proceedings.

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in Strathprints to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profitmaking activities or any commercial gain. You may freely distribute the url (<http://eprints.cdlr.strath.ac.uk>) of the Strathprints website.

Any correspondence concerning this service should be sent to The Strathprints Administrator: eprints@cis.strath.ac.uk

Making science count: Open Access and its impact on the visibility of science

Derek Law
University of Strathclyde,
Glasgow

Introduction

Science publishing is a large global industry. A recent study indicates the sheer volume of what is involved (Worlock,2006). It is believed that there are over twenty thousand peer-reviewed scholarly journals serving a multi-million audience of whom about 5.5 million are themselves researchers. The majority of those journals – some 60% - are now available online, although most also still have print versions. A growing number, currently around 10%, operate some form of open access publication model. This industry generated over €4 billion for English language STM journals alone in 2004. Although that is only a proportion of the total, these are the only available figures which offer reasonably accurate figures on the scale of this huge industry. And it is precisely the sheer size of this global economic phenomenon which blinds many publishers and even authors to the fact that publishing exists to support research; research does not exist to support publishing.

Scholarly Communication is exactly that. From its origins it has been about communicating the results of research both to the peer group and to the wider public. The benefit offered by technology is that we can both begin to look at multiple routes for communication and also begin to measure some of the impacts. Technology has also made informal communication much more important as a part of scientific communication. Of course, citation counts and impact factors are seen as important and have been with us for some time, but to these has been added the impact of downloads from repositories. At present a debate rages over the claim that articles available through open access (OA) are more frequently cited than those which are not freely available. While even the sceptics accept that the deposit of articles in OA repositories seems to be associated with a larger number of citations, and earlier citations for articles, the reasons for this are judged to be less clear, with one view being that authors deposit only their best work in repositories. There is much less consensus over the effect of OA journals, with the evidence seen as patchy and inconsistent. (Worlock, 2006).

The Future of Scholarly Communication

The Internet and the World Wide Web have brought undreamt of opportunities and problems to scholarly communication. And yet the issues remain fundamentally the same as always – Archiving, Access and the Advancement of Science.

Archiving. Hitherto a network of national and university libraries has ensured the retention of the scientific corpus through the provision of copies of published works at different and multiple locations. This infrastructure is much less certain in an electronic environment where information is typically leased rather than purchased and where legal deposit remains unusual. And yet the short life span of publishers compared with the longevity of libraries and universities is almost legendary. Nor has it been the historic role of publishers to ensure permanent archiving. And yet science is built on previous results. It famously exists in the quotation “standing on the shoulders of giants”, a phrase used as the motto of Google Scholar, but traceable back via Isaac Newton to Bernard of Chartres. That long term archiving and preservation role looks to have its best support in repositories.

Access. The PubMed 1000 Exercise (Kiley, 2007) showed that even in the best found libraries and organisations, researchers did not have access to 10-20% of the relevant published papers. For most researchers the position is much worse. If research is to be read and if research is to be seen, it must be readily accessible. Again repositories provide better access free at the point of use than commercial journals with their the toll barriers. That said institutional repositories depend on a vibrant publishing industry which not just allows but encourages self archiving.

Advancement of Science. Repositories have the additional value of promoting both the dissemination of results and academic discourse. Increasingly repositories allow the addition of an enriched range of data to accompany articles, whether underlying data, comment, annotation or links to blogs and wikis. This in turn allows opportunities to take advantage of technology to undertake activities such as data mining and text mining to provide better access to the academic record. As new methods of scientific working emerge, conventional publication becomes only one way of advancing a discipline. Projects such as Neurocommons (www.neurocommons.org) or OpenWetWare (http://openwetware.org/wiki/Main_Page) demonstrate the role of

technology in advancing collaborative working. Repositories then have a key role to play here.

History

The movement to create subject based repositories can be dated to The arXiv, which was originally developed by Paul Ginsparg and started in 1991 as an archive for preprints in physics, but later expanded to include mathematics, computer science, nonlinear science and, most recently, quantitative biology. A small number of other subject based repositories have then developed over the last decade. The trend shifted towards institutional repositories from about the year 2000 as easy to use shareware became widely available, followed in 2001 by the creation of the now widely adopted OAI-PMH protocol which allowed harvesting of data from repositories. A whole series of policy statements have followed, each marking a step in the growing prominence and importance of the Open Access movement:

The first was the Budapest Open Access Initiative of December 2001. There were thirteen initial signatories, a number which had grown to over 360 organizations and 4,000 individuals by August 2006.

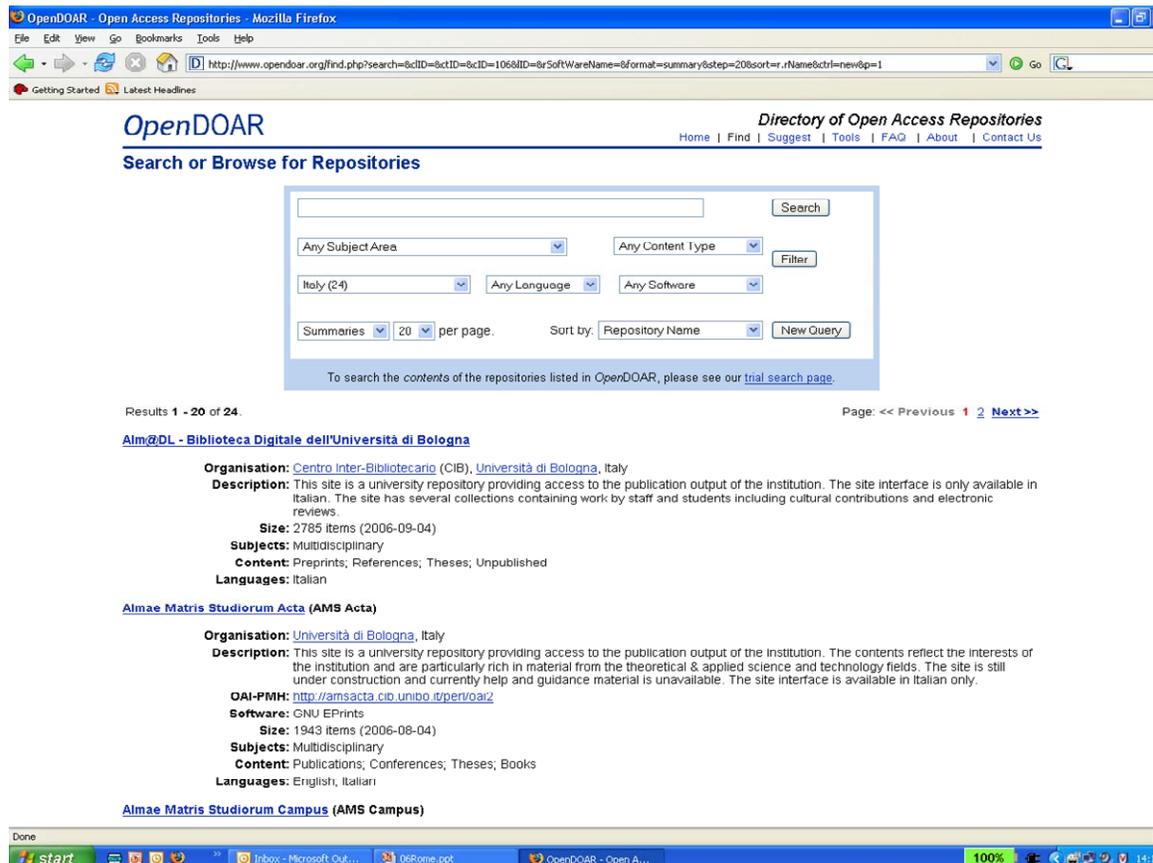
This was followed by a string of national or research funder statements supporting OA

- the Bethesda Statement, 2003
- the Berlin Declaration, 2003
- the Scottish Declaration, 2004
- the National Institutes of Health, 2004
- Research Councils UK, 2006,
- The Bangalore Policy statement for Developing Countries

Significant favourable reports added further momentum from the UK Science and Technology Committee in 2004 and the European Commission in 2006.

All of this activity has led to an organised system of deposit with information shared and managed to common standards worldwide. For example, at a national level the Dutch Cream of Science project highlighting the quality of Dutch science has been hugely influential while services such as OAISTER, ROMEIO and DOAR provide evidence of the mushrooming growth of repositories and of the large number of publishers who are content to see articles deposited by authors. Figure 1 shows a screen from the OpenDOAR site which is a good example of how information is made available on repositories worldwide.

Figure 1

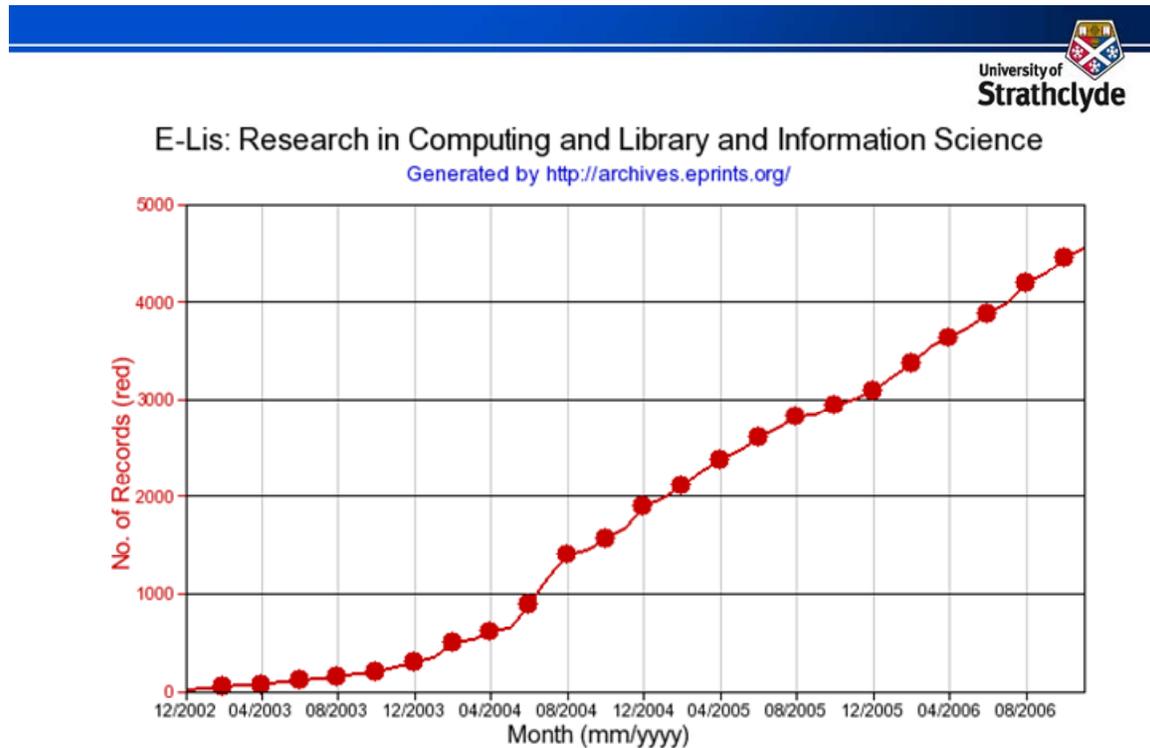


Current Activity

OpenDOAR and OIASTER both show that some eight hundred repositories now exist worldwide, compared with some 250 in 2004. The number continues to grow. Between them, they hold some twelve million articles. There is a widespread feeling that some kind of critical mass has been achieved with these numbers. This is enhanced by a number of large scale national and regional initiatives, for example JISC is funding Higher Education initiatives in the UK; SURF is funding the DARE programme in the Netherlands; the Australian Department for Education Science and Training is funding the ARROW Programme, while the DRIVER Project is an EU initiative. At the same time many publishers are shifting position and there is a plethora of hybrid options, open access journal initiatives and growing permission to self-archive in local repositories. The change seems both rapid and unstoppable.

We can then see that awareness of Open Access is increasing amongst scholars in all disciplines, and while the number of repositories has increased at an average of 1 per day over the last year, the rate of deposit of articles has also increased. This can be demonstrated by this typical graph for the E-Lis subject repository:

Figure 2



Issues

However the need to continue to press to make deposit the norm is equally clear (Swan, 2006). Although there are some 800 repositories globally, there are only 32 documented policies and only 10 mandates – although these numbers are slowly climbing. Worse, only 15% of research articles are spontaneously self-archived, while the average number of postprints self-archived in institutional repositories is a mere 297. This is odd, since there is clear evidence that such self-archiving and the subsequent on-line availability stimulates citation. Since this is increasingly a measure used to assess scientists and their work it is surprising that individuals do not more actively seek to increase the number of citations they receive. Lawrence (2001) has shown a four fold increase in citations in Computer Science, Brody

(2004) has shown a similar four-fold increase in citations in Physics. Mueller (2006) found a similarly large change in impact factors in general internal medicine journals. Hajjem(2006) demonstrated the same effect within biology, business, psychology and sociology journals and Antelman (2005) in philosophy, politics, electrical & electronic engineering and mathematics. There seems overwhelming evidence that such self-archiving increases citation and therefore the effectiveness of scholarly communication in all disciplines.

There is no real barrier to deposit. Some 92% of journals permit self-archiving as shown by the SHERPA/RoMEO site at <http://romeo.eprints.org/stats.php> or www.sherpa.ac.uk/romeo.php

Yet Swan (2006) has discovered that only 24% of authors have deposited papers in a repository, while only 15% of researchers deposit regularly in institutional archives. This figure will surely change as the major funding agencies mandate deposit. These figures are all the more surprising given that such deposit not only impacts on citation volume but also impacts on citation speed – and hence influences recognition of the research.

Swan (2006a) has demonstrated that Open Access articles are cited earlier, and, as shown below, they are downloaded more often. As Figure 3 (below) demonstrates, open access abstracts are viewed more frequently. As Swan's work has shown there is a significant correlation between downloads today and citations two years later. This correlation has two immediate implications:

(1) Download counts can be used as early performance indicators for papers and authors, even before their impact is reflected in citation counts

(2) Enhancing usage impact is yet another reason for authors to provide open access to their articles by self-archiving them.

As a small example take a paper by this author describing convergence of support services at the University of Strathclyde. This was published in the United States as a book chapter late in 2003. It has not yet been cited (to the author's knowledge) and does not appear in Google Scholar. It was mounted in the institutional repository in late 2005 and over twelve months attracted attention as follows:

Figure3

[The place of useful learning: convergence at the University of Strathclyde](#)

For this eprint: [[Past four weeks](#)] [[This year](#)] [[Last year](#)] [[All years](#)]

Abstract views and document downloads for all years

The numbers in (parentheses) are the number of distinct countries that views/downloads originated from.

	Abstracts	Downloads
Views	288 (15)	84 (7)

Views by country (derived from IP address of query) for all years

Country	Abstracts	Downloads
 United States	215	50
 Our Intranet	32	13
 United Kingdom	13	11
 Switzerland	4	6
 Czech Republic	3	2
 Ireland	1	1
 Palestinian Territory	1	1
 Japan	6	0
 Germany	3	0
 Italy	3	0
 Sweden	3	0
 India	1	0
 Spain	1	0
 Poland	1	0
 Portugal	1	0
Grand Totals:	288	84

abstract views originating from 15 distinct countries

document downloads originating from 7 distinct countries

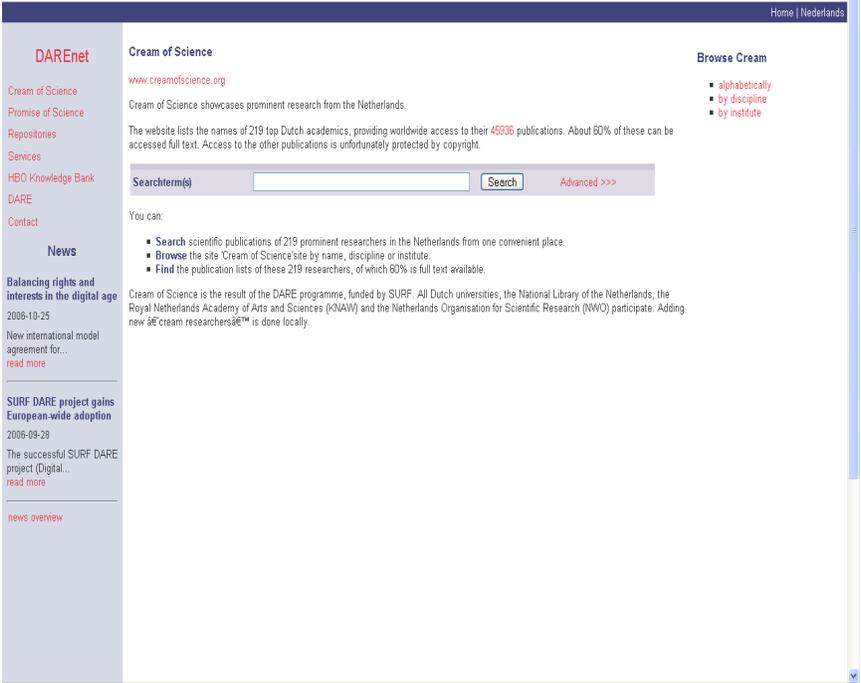
What is interesting is not just the interest generated by deposit, but the spread of countries from which searches have been made. This is evidently much greater than the reach of an expensive monograph from a small American publishing house.

Now whatever view is taken of research quality there are a variety of models which attempt to express it through metrics. Impact factors, h-factors, g-factors and the rest are all used in an attempt to balance quality and quantity (Lehmann, 2006). What is common to all of them is the use of citations to assess impact. When aggregated these are used to create league tables of institutional quality. Ultimately this is a factor in the award of institutional funding. There is then every reason for researchers to see an incentive to use repositories to increase citations.

Other initiatives

Institutional Repositories are not, of course, the only vehicle for promoting the outputs of science. Perhaps the best known national initiative is Cream of Science in the Netherlands (DARE, 2007).

Figure 4.



The screenshot shows the Cream of Science website interface. At the top right, there is a logo for the University of Strathclyde. The main content area is titled 'Cream of Science' and includes the URL www.creamofscience.org. Below the title, there is a search bar with the placeholder text 'Searchterm(s)', a 'Search' button, and a link to 'Advanced >>'. To the right of the search bar, there are three options for browsing: 'Browse Cream', 'alphabetically', 'by discipline', and 'by institute'. The page also features a 'DAREnet' sidebar with various links and a 'News' section with several articles.

This project has identified the principle scientists in that country and then sought to deposit the work of 229 of them in a repository to showcase the best research. This has obvious resonance with a variety of government agendas to promote the country in every sphere and appeals to the personal vanity of the individual. This initiative has been much admired and is already being explored as a model for the UK.

Another major initiative has just been launched in medicine. Although it's initial focus and funding is from the UK, it seems to have clear European ambitions. Based on the well known PubMed Central - the US National Institutes of Health free digital archive of biomedical and life sciences journal literature – UK PubMed Central aims to provides a stable, permanent and free-to-access online digital archive of full-text, peer-reviewed research publications.

Launched in January 2007, the initial phase of developing UKPMC involves mirroring the PMC database, and implementing a manuscript submission system to enable UK scientists to submit their research papers for inclusion in UKPMC. The project is supported by the eight major UK biomedical funding agencies who between them fund over 90% of research in biomedicine in the UK. The fact that they have mandated deposit will not only lead to a much higher rate of deposit but should give UK scientists an edge in terms of citation.

Figure 5. UKPubMedCentral

UKPubMedCentral

University of Strathclyde

wellcome trust

About us

Funding Strategy Funded activities Knowledge centre What's on About us

Home > Strategy > Policies and advocacy > Spotlight issues > Open access

Strategic Plan 2005 - 2010

Policies and advocacy

Policies and positions

Spotlight issues

Submissions to consultations

OPEN AND UNRESTRICTED ACCESS TO THE OUTPUTS OF PUBLISHED RESEARCH

Policy summary

It is a fundamental part of our charitable mission to ensure that the work we fund can be read and utilised by the widest possible audience. In support of this objective the Wellcome Trust [Grant Conditions](#) have been amended such that electronic copies of any research papers that have been accepted for publication in a peer-reviewed journal, and are supported in whole or in part by Wellcome Trust funding, must be deposited into PubMed Central (PMC) (or UKPMC once established). These papers will be made freely available via these web-based repositories as soon as possible, and in any event within six months of the journal publishers' official date of final publication. This requirement applies to all grants, regardless of award date. For further details, see the full [position statement](#) on open and unrestricted access to the outputs of published research, and the [authors' guide and FAQ](#).

Latest developments: September 2006

The Wellcome Trust, as part of a group of UK research funders [PDF 24KB], has announced that the contract to run UK PubMed Central (UKPMC) has been awarded to a partnership between the British Library, the University of Manchester and the European Bioinformatics Institute (EBI).

Based on the PMC model currently used by the US National Library of Medicine at the National Institutes of Health, UKPMC will provide **free online access** to the digital archive of published articles resulting from research paid for by any of the funding consortium. It will be **fully searchable** and extensively linked to other online resources.

Launch of the service is scheduled for **January 2007**. See the [press release](#) (31 July 2006) for further information.

Page last updated: 20 September 2006

Gibbs Building, 215 Euston Road, London NW1 2BE, UK. tel: +44 (0)20 7611 8888 email: contact@wellcome.ac.uk

Why does it matter?

Increasingly metrics are being used to determine the quality of research. Increasingly we can expect citation, downloads etc. to determine not only past performance but also future grants. Already free software exists which allows a direct comparison of the impact of individuals (<http://www.harzing.com/resources.htm>). However many reservations one may have and how ever crude a tool this might be, it will be used. One of the most systematised structures operates in the United Kingdom, where individual researchers are assessed into five categories

- 0 = not submitted for assessment
- 1* = nationally significant
- 2* = international reputation
- 3* = working ONLY at international level
- 4* = global superstar

A popular prejudice is that such 4* "galacticos" are probably Nobel prizewinners or at least have their own television series, while "normal researchers will be 2* or 3*. Most of the judgements are based on metrics, which are measured by peer group esteem; the research environment; publications and increasingly their citations.

Increasingly too, universities wish to perform well in the published league tables of universities. If we look at the world league table of universities, it uses only two metrics:

- Who knows/mentions the institution in a survey
- Citation count

Universities wish to be seen to be well placed in these league tables. And how are these measures most easily increased?.... By depositing papers in the institutional repository!

As Sir John Sulston, Nobel Prizewinner and the British scientist behind the Human Genome Project put it "Ensuring that the outputs of research are freely available to all is the best way to maximise their utility" (Kiley, 2007)

References

Antelman, K. (2004) Do Open-Access Articles Have a Greater Research Impact? *College & Research Libraries* 65 pp372-382

Brody, T., Stamerjohanns, H., Vallieres, F., Harnad, S., Yves, G. and Charles, O. (2004) The effect of Open Access on Citation Impact.
<http://eprints.resist.ecs.soton.ac.uk/9941/>

DARE (2007) <http://www.creamofscience.org/en/page/language.view/keur.page>
[page viewed on 19th February 2007]

Hajjem, C., Gingras, Y., Brody, T., Carr, L. and Harnad, S. (2005) Open Access to Research Increases Citation Impact. Technical Report, Institut des sciences cognitives, Université du Québec à Montréal.

Kiley, Robert (2007) cited by him in a presentation made to the 2007 Conference *Scientific Publishing in the European Research Area* Brussels, 2007

Lawrence S. Online or Invisible? *Nature* (2001) 411:521.
<http://www.neci.nec.com/~lawrence/papers/online-nature01/>

Lehmann, S., Jackson, Andrew D., Lautrup, Benny E. (2006). Measures for Measures *Nature* 444 pp1003-1004

Mueller, P.S., Murali, N.S., Cha, S.S., Erwin, P.J., Ghosh, A.K. The effect of online status on the impact factors of general internal medicine journals *Netherlands journal of medicine* (2006), Vol. 64, No. 2

Swan, Alma (2006) *Repositories overview: Policies and implementation*. Powerpoint presentation given at the Open Scholarship 2006 Conference.
<http://hdl.handle.net/1905/649>

Swan, Alma (2006a) The institutional repository: what it can do for your institution and what the institution can do for the repository. ANKOS Workshop 2006

Worlock, David *Assessing the Evidence*. Presentation to the STM Frankfurt Conference in 2006 reporting on the "UK scholarly journals: 2006 baseline report commissioned by the Research Information Network.