Main Articles

Mandated online RAE CVs linked to university eprint archives:

Enhancing UK research impact and assessment

Stevan Harnad, Les Carr, Tim Brody and **Charles Oppenheim** make a compelling case for optimising the UK's
pre-eminence in Research Assessment.

| Main Contents | Section Menu | Email Ariadne | Search Ariadne |

Introduction

Being the only country with a national research assessment exercise [1], the UK is today in a unique position to make a small change that will confer some large benefits. The Funding Councils should mandate that in order to be eligible for Research Assessment and funding, all UK research-active university staff must maintain (I) a standardised online RAE-CV, including all designated RAE performance indicators, chief among them being (II) the full text of every refereed research paper, publicly self-archived in the university's online Eprint Archive and linked to the CV for online harvesting, scientometric analysis and assessment. This will (i) give the UK Research Assessment Exercise (RAE) far richer, more sensitive and more predictive measures of research productivity and impact, for far less cost and effort (both to the RAE and to the universities preparing their RAE submissions), (ii) increase the uptake and impact of UK research output, by increasing its visibility, accessibility and usage, and (iii) set an example for the rest of the world that will almost certainly be emulated, in both respects: research assessment and research access.
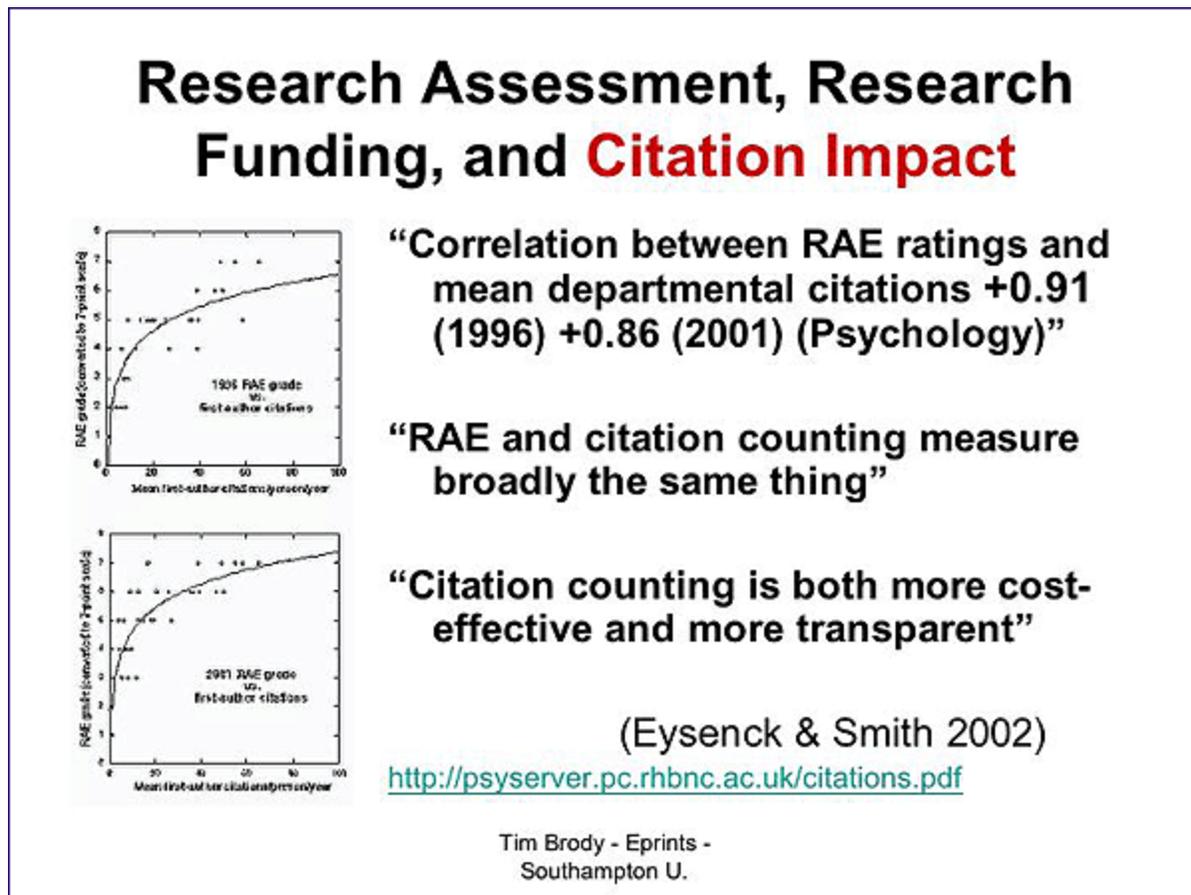
Figure 1: Predicting RAE Ratings from Citation Impact (Smith & Eysenck 2002)

The UK already has a Research Assessment Exercise (RAE), every 4 years. The RAE costs a great deal of time and energy to prepare and assess, for both universities and assessors (time and energy that could be better used to actually do research, rather than preparing and assessing RAE returns).

The Journal Impact Factor

An important and predictive measure of research impact in many areas of research is the Journal Impact Factor (JIF) of the journal in which the article appears. The JIF is the average number of citations per article per year. The number of times an article has been cited (hence used) is an indication of the importance and uptake of that research. For core journals in all subject areas the JIF can be obtained from the Institute of Scientific Information's Journal Citation Reports service, for which the UK has a national site license [2].
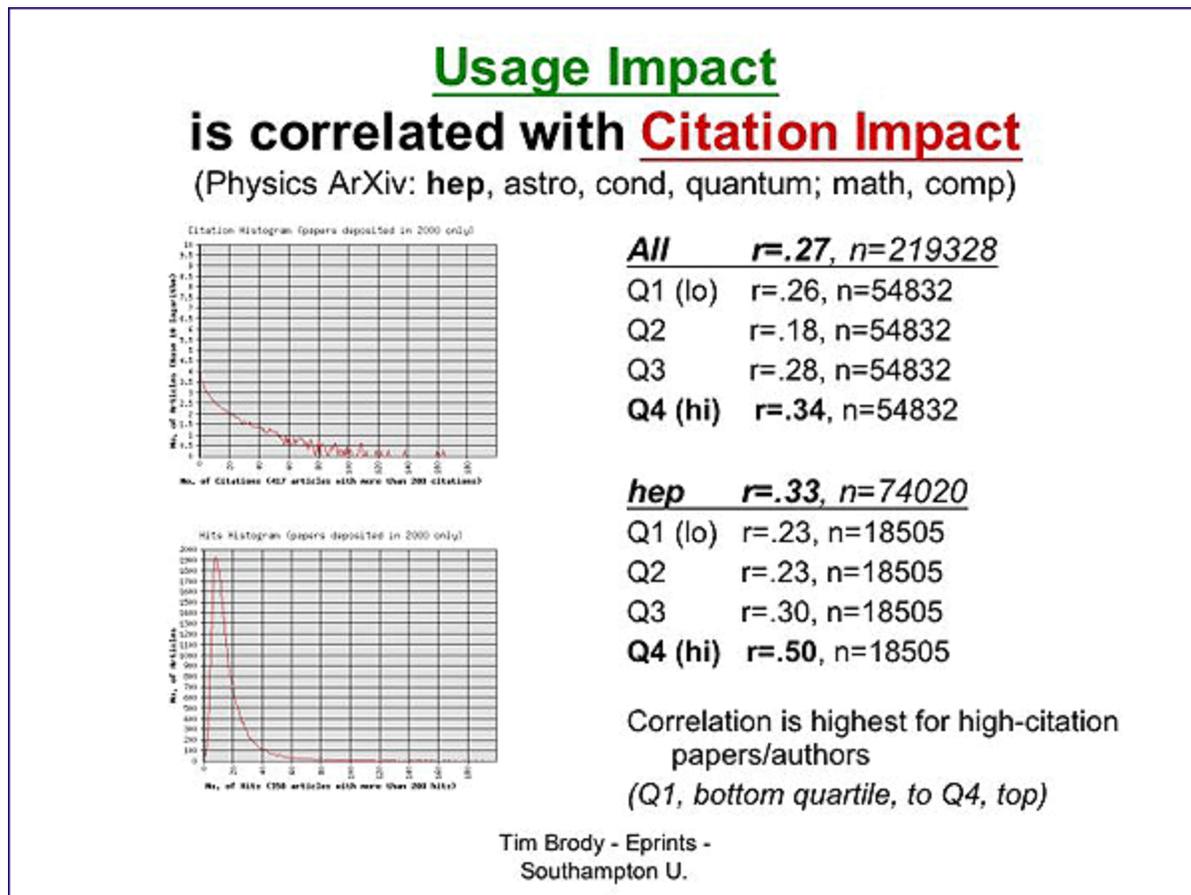
Figure 2: Predicting Citation Impact From Usage Impact (Physics ArXiv)

The JIF figures only indirectly in the RAE: researchers currently have to submit 4 publications for the 4-year interval. It is no secret that departments (informally) weight candidate papers by their JIFs in deciding on what and whom to submit. Although it is always stressed by the RAE panels that they will **not** judge papers by the journals in which they appeared (but by the quality of their content), it would nevertheless be a strange RAE reviewer who was indifferent to the track record, refereeing standards, and rejection rate of the journal whose quality-standards a paper has met. (For books or other kinds of publications, see below; in general, peer-reviewed journal or conference papers are the coin of the research realm, especially in scientific disciplines.)

Statistical correlational analyses on the numerical outcome of the RAE using average citation frequencies predict departmental outcome ratings remarkably closely. Smith & Eysenck [3], for example, found a correlation of as high as .91 in Psychology (Figure 1). Oppenheim and collaborators [4] [5] and Holmes & Oppenheim [6] found correlations of .80 and higher in other disciplines.

The power of the indirect journal-based JIF has not yet been tested for predicting RAE rankings, but it is no doubt correlated with the well demonstrated RAE predictive power of direct author-based citation counts (average or total). Journal impact is the blunter instrument, author or paper impact the sharper one (Seglen [7]). But a natural conclusion is that the reliability and validity of RAE rankings can and should be maximised by adding and testing as many candidate predictors as possible to a weighted multiple regression equation.

Some old and new scientometric ("publish or perish") indices of research impact
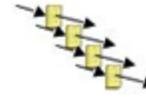
- quality-level and citation-counts of the <u>journal</u> in which the article appears
- citation-counts for the <u>article</u>
- citation-counts for the <u>researcher</u>
- co-citations, co-text (cited <u>with whom/what else</u>?)
- citation-counts for the <u>preprint</u>
- usage-measures ("<u>hits</u>," webmetrics)
- <u>time-course</u> analyses, early predictors, etc. etc.

Tim Brody - Eprints -
Southampton U.

Figure 3: New Online Performance Indicators

Nor is there any reason why the RAE should be done, at great effort and expense, every 4 years! Since the main determining factor in the RAE outcome ratings is research impact, there is no reason why research impact should not be **continuously** assessed, using not only author and paper citation counts and the JIF, but the many other measures derivable from such a rich research performance indicator database. There is now not only a method to assess UK research impact (i) continuously, (ii) far more cheaply and effortlessly for all involved, and (iii) far more sensitively and accurately (Figures 2-4), but doing the RAE this new way will also dramatically enhance UK research impact itself, (iv) increasing research visibility, usage, citation and productivity, simply by maximising its accessibility (Figures 5-8).
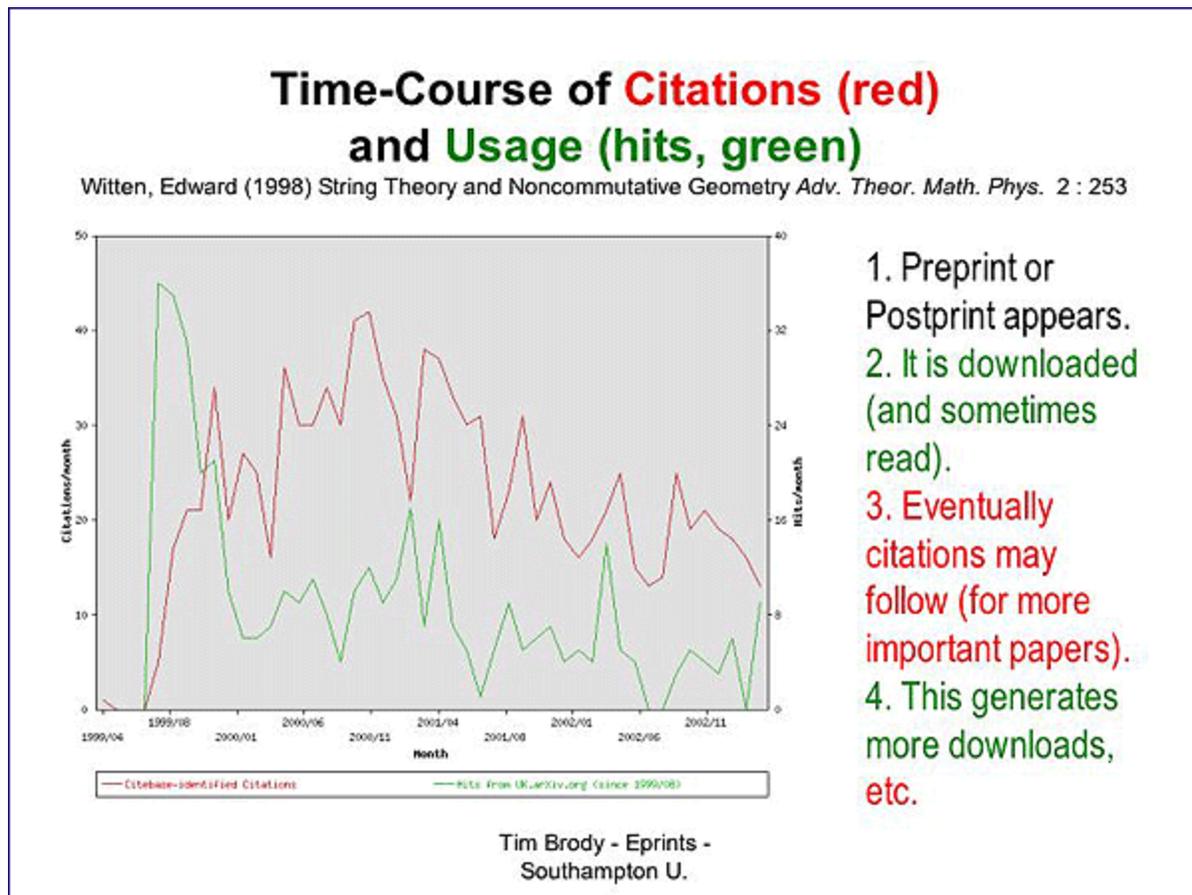
Figure 4: Time-Course of Citations and Usage (Physics ArXiv)

RAE Online

The method in question is to implement the RAE henceforth online-only, with only two critical components: (a) a continuously updated and continuously accessible RAE-standardised online CV (containing all potential performance indicators: publications, grants, doctoral students, presentations, etc.) [8] for every researcher plus (b) a link from each CV to the full digital text of every published paper _ books discussed separately below _ self-archived in that researcher's university Eprint Archive (an online archive of that institution's peer-reviewed research output) [9]. (See the free, open-source software developed at Southampton to allow universities to create their own institutional Eprint Archives [10]).
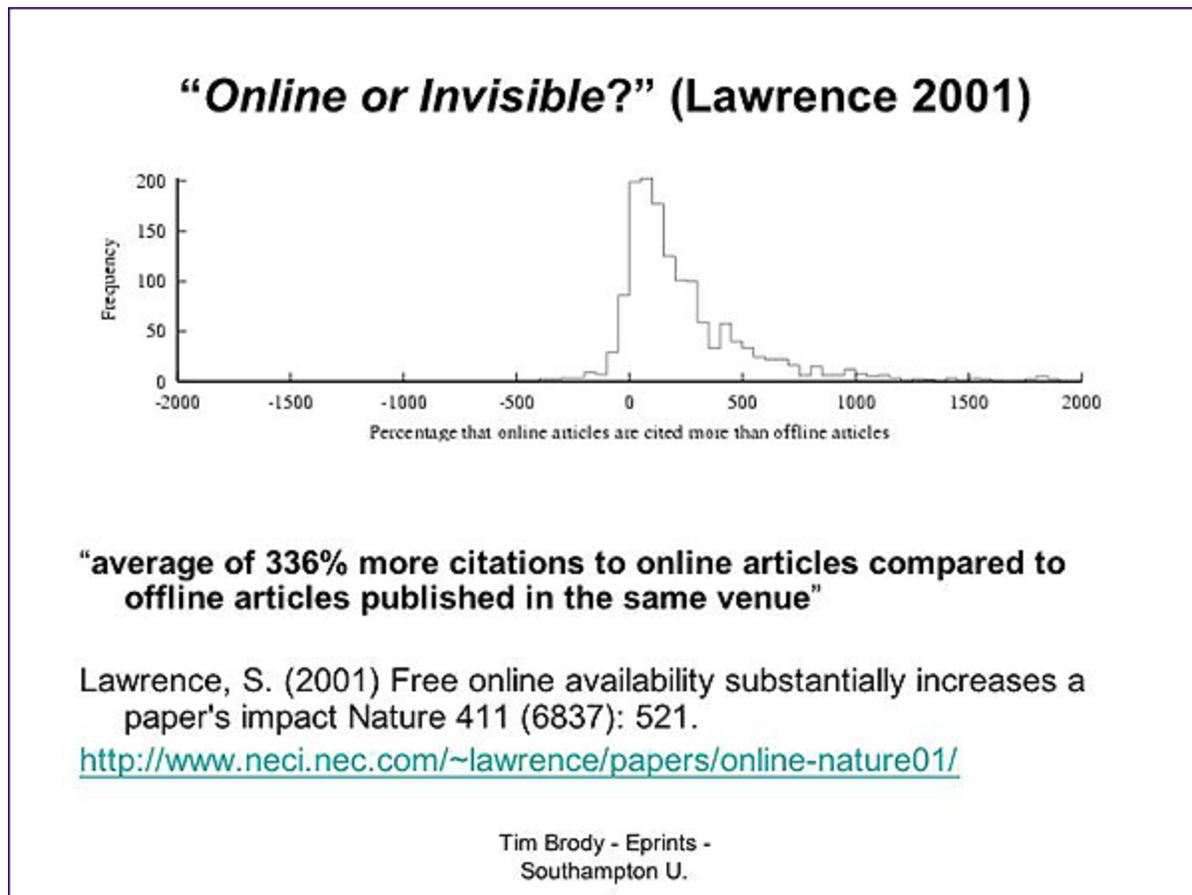
Figure 5: Open Online Full-Text Access Enhances Citations Dramatically (Computer Science)

Currently, university peer-reviewed research output _ funded by government research grants, the researcher's time paid for by the researcher's institution _ is given, free, by all researchers, to the peer-reviewed journals in which it appears. The peer-reviewed journals in turn perform the peer-review, which assesses, improves and then certifies the quality of the research; (this is one of the indirect reasons why the RAE depends on peer-reviewed journal publications) (Figure 7). There is a hierarchy of peer-reviewed journals, from those with the highest quality standards (and hence usually the highest rejection rates and impact factors) at the top, grading all the way down to the lowest-quality journals at the bottom [11]. The peers review for free; they are just the researchers again, wearing other hats. But it costs the journals something to implement the peer reviewing [12].

Research Impact

I.   measures the <u>size</u> of a research contribution to further research

II.   generates further research <u>funding</u>

III. contributes to the research <u>productivity</u> and financial support of the researcher's <u>institution</u>

IV.   advances the researcher's <u>career</u>

V.   promotes research <u>progress</u>

Tim Brody - Eprints - Southampton U.

Figure 6: The Vast and Varied Influence of Research Impact

Partly because of the cost of peer review, but mostly because of the much larger cost of print on paper and its dissemination, plus online enhancements, journals make charges (subscriptions, licenses, pay-per-view) for access to researchers' papers, (even though the researchers gave them the papers for free [13]). The effect is a great loss of potential research impact, because most institutions cannot afford to pay the access-tolls to most peer-reviewed journals, (there are 20,000 in all, across disciplines), but only to a small and shrinking proportion of them [14].

Opening Access

Hence the second dramatic effect of revising the RAE -- so as to transform it into online continuous assessment based on the institutional self-archiving of all UK peer-reviewed research output -- will be to make all that UK research accessible to all those would-be users worldwide whose access is currently blocked by toll barriers (Figure 8). If RAE mandates self-archiving, university departments will mandate it too. See, for example, the draft Southampton self-archiving policy [15].

The UK full-text peer-reviewed research archives will not only be continuously accessible to all potential users, but the access will be continuously assessable, in the form not only of continuously updated impact estimates based on the classical measure of impact, i.e. citations, but usage will also be measured at earlier stages than citation, namely downloads ("hits," Figure 2)[16] of both peer-reviewed "postprints" and pre-refereeing "preprints". Many powerful new online measures of research productivity and impact will also develop around this rich UK research performance database (Figures 3-4), further increasing the sensitivity and predictiveness of the RAE analyses. (See the online impact-measuring scientometric search engines we have developed at Southampton [17] [18]).

All that is required (Figure 9) is for the RAE to move to online submissions, mandating online CVs linked to the full-text draft of each peer-reviewed publication in the researcher's institutional Eprint Archive [19]. Reference link-based impact assessment engines like citebase and Web of Science [20] can then be used by RAE to derive ever richer and more accurate measures of research productivity and impact (Figure 3), available to the RAE continuously. Universities could continuously monitor and improve their own research productivity and impact, using those same measures. And the rest of the world could see and emulate the system, and its measurable effects on research visibility, uptake and impact.

What to Do about Royalty-Bearing Books?

Just a few loose ends: books are usually not give-aways, as peer-reviewed research is, so full-text self-archiving is probably not viable for book output, (apart from esoteric monographs that produce virtually no royalty revenue). But even if the book's full text itself cannot be made accessible online, its metadata and references can be! Then the citation of books by the online peer-reviewed publications becomes a measurable and usable estimate of their impact too. And for disciplines where research and productivity do not consist of text at all, but of other forms of digital output, both online usage counts and citations by text publications can still be used to estimate impact; and there are always the further kinds of performance indicators in the standardised RAE-CV that can be used to design discipline-specific metrics.

Conclusions

The UK is uniquely placed to move ahead with this and lead the world, because the RAE is already in place. The Netherlands has no formal RAE yet, but it is about to implement a national system of open research archiving for all of its universities called DARE [21]. It is just a matter of time before they too realise that a marriage between a national network of DARE-style institutional Eprint Archives and CVs plus a national RAE-style research assessment exercise make a natural, indeed an optimal combination.

But although the naturalness and optimality _ indeed the inevitability _ of all this is quite transparent, it is a fact that research culture is slow to change of its own accord, even in what is in its own best interests. That, however, is precisely why we have funding councils and research assessment: to make sure that researchers do what is best for themselves, and best for research, and hence also best for the supporters (and beneficiaries) of research, namely, tax-paying society. The institutional self-archiving of research output, for the sake of maximising research access and impact, has been much too slow in coming, even though it has already been within reach for several years. The UK and the RAE are now in a position to lead the world research community to the optimal and the inevitable [22].

We at Southampton and Loughborough, meanwhile, keep trying to do our bit to hasten the optimal/inevitable for research and researchers. At Loughborough we are clearing the way for universal self-archiving of university research output by sorting out the copyright issues (and non-issues [23]). At Southampton we are planning to harvest all the metadata form the submissions to RAE 2001 [24] into RAEprints, a "meta-archive" that is intended to demonstrate what RAE returns would look like if this RAE upgrade proposal were adopted. Of course (i) RAEprints will contain only four papers per researcher, rather than their full peer-reviewed research output, (ii) it will only contain the metadata for those papers (author, title, journal name), not the full text and the all-important references cited. But we will also try to enhance the demo by adding as much of this missing data as we can find – both from Journal Citation Reports [20] and from the Web itself – to give at least a taste of the possibilities: using paracite [25] an on-line citation-seeker that goes out and tries to find peer-reviewed full-text papers on the Web, we will "stock" RAEprints with as much as

turns out to be available online currently _ and then we will invite all the RAE 2001 researcher/authors to add their full texts to RAEprints too!

We hope that the UK Funding Councils will put their full weight behind our recommended approach (Figure 9) when they publish their long-awaited review of the RAE process [26].
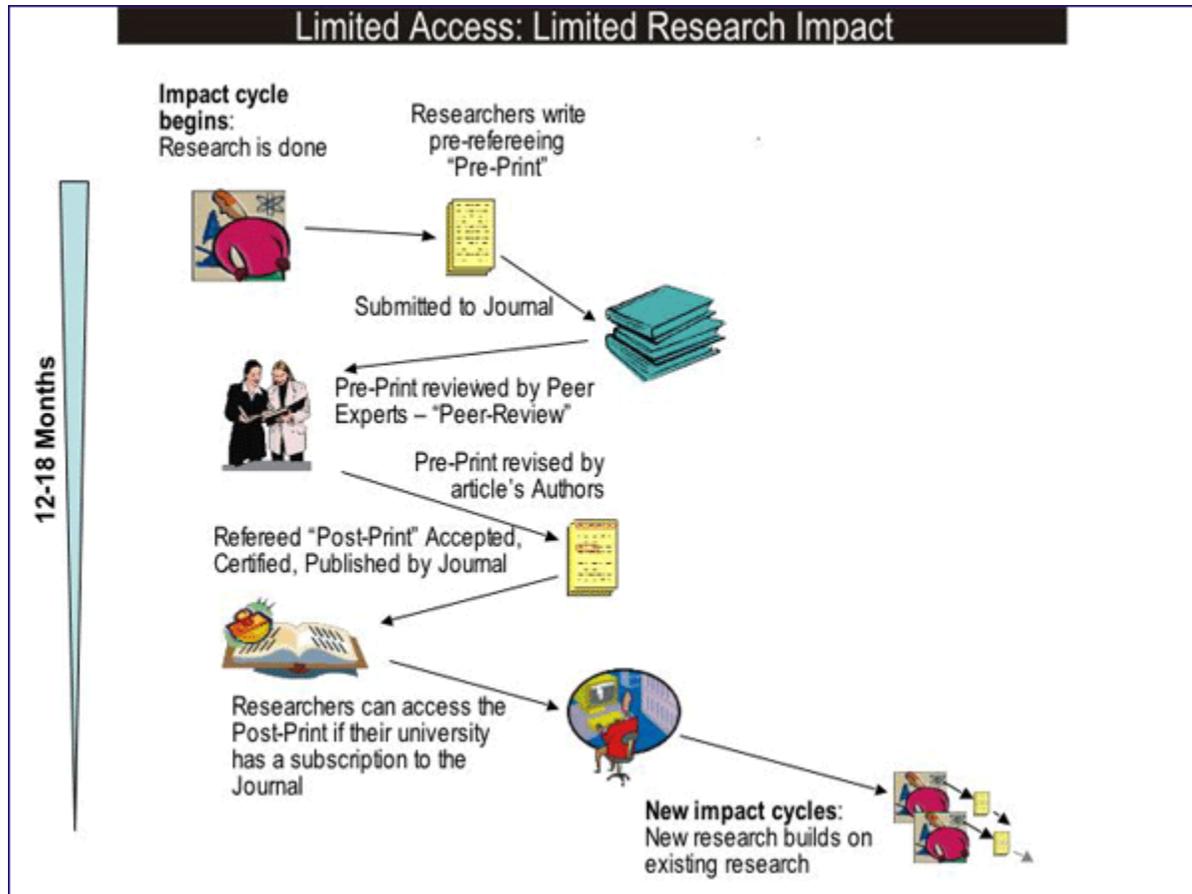

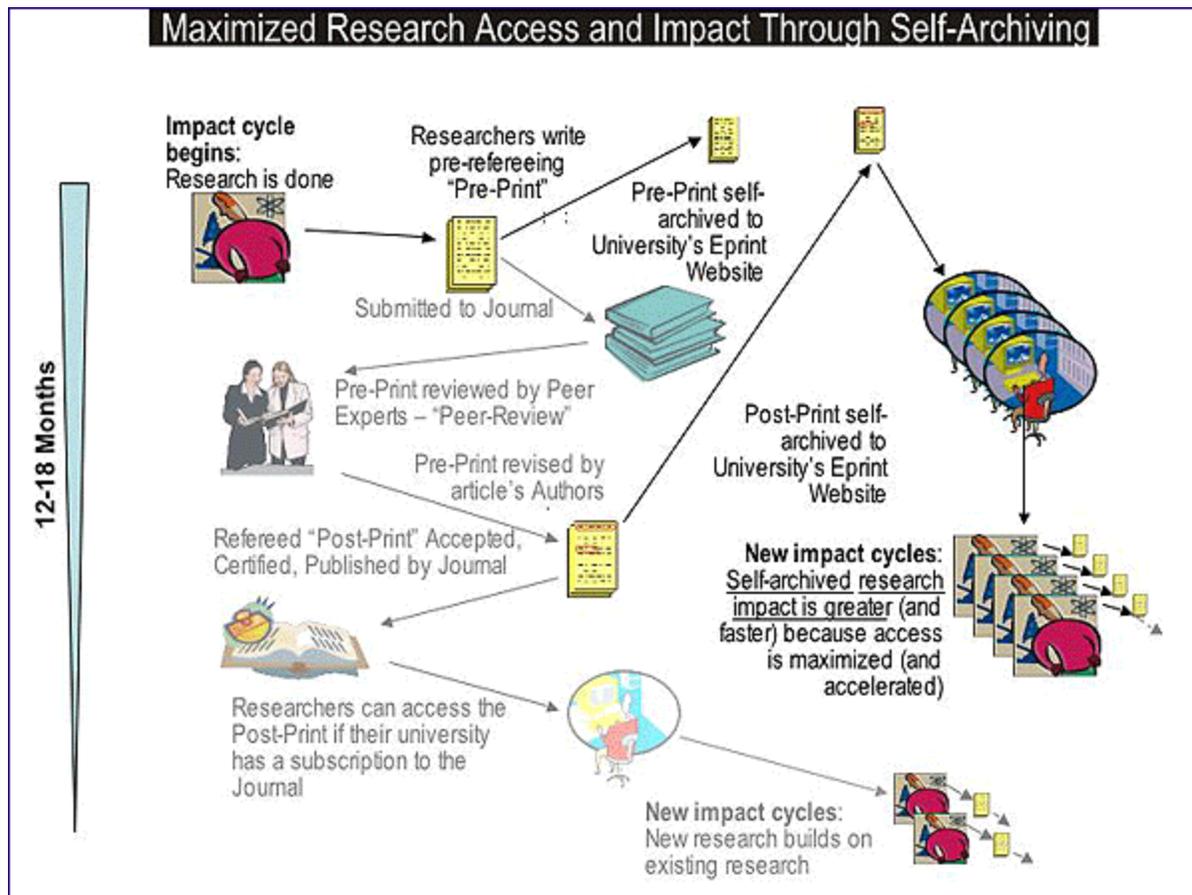**Figure 7: The Limited Impact Provided By Toll-Based Access Alone**

Figure 8: Maximising Research Impact Through Self-Archiving of University Research Output

What is needed to fill the archives:

1. *Universities*: Adopt a university-wide policy of self-archiving all university research output
2. *Departments*: Create Departmental OAI-compliant Eprint Archives
3. *University Libraries*: Provide digital library support for research self-archiving and archive-maintenance
4. *Promotion Committees*: Request a standardized online CV from all candidates, with refereed publications all linked to their full-texts in the Departmental Archives
5. *Research Funders*: Assess research impact online (from the online CVs)

Tim Brody - Eprints -
Southampton U.

Figure 9: What Needs to be Done to Fill the Eprint Archives

References

1. 2001 RAE Submissions http://www.hero.ac.uk/rae/submissions/
2. Journal Citation Reports Service for UK Education http://wos.mimas.ac.uk/jcrweb/
3. Smith, Andrew, & Eysenck, Michael (2002) "The correlation between RAE ratings and citation counts in psychology," June 2002 http://psyserver.pc.rhbnc.ac.uk/citations.pdf
4. Oppenheim, Charles (1995) The correlation between citation counts and the 1992 Research Assessment Exercises ratings for British library and information science departments, Journal of Documentation, 51:18-27.
5. Oppenheim, Charles (1998) The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology, Journal of Documentation, 53:477-87.
   http://dois.mimas.ac.uk/DoIS/data/Articles/julkokltny:1998:v:54:i:5:p:477-487.html
6. Holmes, Alison & Oppenheim, Charles (2001) Use of citation analysis to predict the outcome of the 2001 Research Assessment Exercise for Unit of Assessment (UoA) 61: Library and Information Management. http://www.shef.ac.uk/~is/publications/infres/paper103.html
7. Seglen, P. O. (1992). The skewness of science. Journal of the American Society for Information Science, 43, 628-638 http://www3.interscience.wiley.com/cgi-bin/abstract/10049716/START
8. See (Draft Demo) Template for UK Standardized CV for Research Assessment (sample data only) http://paracite.eprints.org/cgi-bin/rae_front.cgi
9. How can an institution facilitate the filling of its Eprint Archives?
   http://www.eprints.org/self-faq/#institution-facilitate-filling
10. GNU EPrints 2 http://software.eprints.org/

11. Journal Citation Reports Service for UK Education http://wos.mimas.ac.uk/jcrweb/
12. Peer-review Discussion Group Summary, 21 March 2001
    http://agenda.cern.ch/askArchive.php?base=agenda&categ=a01193&id=a01193s5t11%2Ftransparencies%2F%2FDoyle-peer-review.pdf
13. Harnad, S. (2001) For Whom the Gate Tolls?
    http://www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm
14. ARL Statistics and Management Program http://www.arl.org/stats/index.html
15. ECS Research Self-Archiving Policy http://www.ecs.soton.ac.uk/~lac/archpol.html
16. Both usage statistics and citation statistics are open to potential abuse. See:
    Impact Health Warning to Authors and Users http://citebase.eprints.org/help/#impactwarning
    and
    Potential Metric Abuses (and their Potential Metric Antidotes) (mail)
    http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/2642.html .
    The data nevertheless still have considerable signal value, its effect-size can be estimated, and ways of detecting and correcting for abuses will evolve as these new measures become more important.
17. citebase Search http://citebase.eprints.org/cgi-bin/search
18. The Open Citation Project http://opcit.eprints.org
19. What can research funders do to facilitate self-archiving? http://www.eprints.org/self-faq/#research-funders-do
20. ISI Web of Science Service for UK Education http://wos.mimas.ac.uk/
21. DARE announcement http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/2356.html
22. Self-Archive Unto Others http://www.ecs.soton.ac.uk/~harnad/Temp/unto-others.html
23. Project RoMEO http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html
24. 2001 RAE Submissions http://www.hero.ac.uk/rae/submissions/
25. ParaCite experimental service http://paracite.eprints.org/
26. RA Review http://www.ra-review.ac.uk/

Further Reading

1. Berners-Lee, Tim. & Hendler, Jim. (2001). Scientific publishing on the "semantic web." Nature, 410, 1023-1024
2. Garfield, Eugene (1979). Citation indexing: Its theory and applications in science, technology and the humanities. New York. Wiley lnterscience
3. Harnad, S. (2001) The Self-Archiving Initiative. Nature 410: 1024-1025
   http://www.nature.com/nature/debates/e-access/Articles/harnad.html
4. Harnad, Stevan, & Carr, Les (2000). Integrating, navigating, and analysing open eprint archives through open citation linking (the OpCit project). Current Science. 79(5). 629-638
5. Harnad, Stevan. (2001) Research access, impact and assessment. Times Higher Education Supplement 1487: p. 16. http://www.ecs.soton.ac.uk/~harnad/Tp/thes1.html
6. Harnad, S. (2003) Electronic Preprints and Postprints. Encyclopedia of Library and Information Science Marcel Dekker, Inc. http://www.ecs.soton.ac.uk/~harnad/Temp/eprints.htm
7. Harnad, S. (2003) Online Archives for Peer-Reviewed Journal Publications. International Encyclopedia of Library and Information Science. John Feather & Paul Sturges (eds). Routledge http://www.ecs.soton.ac.uk/~harnad/Temp/archives.htm
8. Hodges, S., Hodges, B., Meadows, A.J., Beaulieu, M. and Law, D. (1996) The use of an algorithmic approach for the assessment of research quality, Scientometrics, 35, 3-13.
9. Jaffe, Sam (2002) Citing UK Science Quality: The next Research Assessment Exercise will probably include citation analysis. The Scientist 16 (22), 54, Nov. 11, 2002. http://www.the-scientist.com/yr2002/nov/prof1_021111.html

10. Ingwersen, P., Larsen, B. and Wormell, I. (2000) Applying diachronic citation analysis to ongoing research program evaluations. In: The Web of Knowledge : a Festschrift in Honor of Eugene Garfield / Cronin, B. & Atkins, H. B. (eds.). Medford, NJ: Information Today Inc. & The American Society for Information Science.
11. Oppenheim, Charles (1996) Do citations count? Citation indexing and the research assessment exercise, Serials, 9:155-61, 1996.
12. Warner, Julian (2000) Research Assessment and Citation Analysis. The Scientist 14(21), 39, Oct. 30, 2000. http://www.the-scientist.com/yr2000/oct/opin_001030.html
13. Thelwall, Mike (2001) Extracting macroscopic information from Web links. Journal of the American Society for Information Science 52(13) (November 2001) http://portal.acm.org/
14. Zhu, J., Meadows, A.J. & Mason, G. (1991) Citations and departmental research ratings. Scientometrics, 21, 171-179

Author Details

Prof. Stevan Harnad
Professor of Cognitive Science
Southampton University
E-mail: harnad@ecs.soton.ac.uk
Website: http://www.ecs.soton.ac.uk/~harnad/

Dr. Les Carr
Senior Lecturer in Electronics and Computer Science
Southampton University
E-mail: lac@ecs.soton.ac.uk
Website: http://www.ecs.soton.ac.uk/~lac/

Tim Brody
Doctoral Candidate in Electronics and Computer Science
Southampton University
E-mail: tdb01r@ecs.soton.ac.uk
Website: http://citebase.eprints.org/

Prof. Charles Oppenheim
Professor of Information Science
Loughborough University
E-mail: C.Oppenheim@lboro.ac.uk
Website:
http://www.lboro.ac.uk/departments/ls/staff/coppenheim.html

Originating URL: http://www.ariadne.ac.uk/issue35/harnad/intro.htm
Copyright and citation information **File last modified**: Friday, 30-Jul-2004 14:26:07 UTC

Main Contents   Section Menu   Email Ariadne   Search Ariadne