Research Dissemination and Invocation on the Web

Mike Thelwall[1]
School of Computing and Information Technology, University of Wolverhampton,
35/49 Lichfield Street, Wolverhampton, WV1 1EQ, UK.
Email: mthelwall@wlv.ac.uk

Abstract
The importance of the Web as a new medium for disseminating and promoting scholarly research is discussed. Particular concern is given to its potential use to provide evidence of wider impact for research than that which can be shown by citation analysis. Recommendations are made for basic strategies for the reporting of the online impact of research leading to the production of what is termed a Web Invocation Portfolio. A conceptual framework is also proposed to help funding and promotion committees assess and compare portfolios.

Introduction

The Web is an important part of research and education in many parts of the world. It is now widely used as one of the primary means of disseminating research findings through digital libraries and electronic documents such as e-journals (Harter and Ford, 2000; Halliday and Oppenheim, 2001), e-print archives (Harnad and Carr, 2000; Garnet et al, 2002; Town *et al*., 2002) and online conference proceedings (Goodrum *et al*., 2001). One recent study found that the online publication of papers in computer science may facilitate higher citation-based impact (Lawrence, 2001). Moreover, according to Weigold (2001), "[the Web] has the potential to dramatically change the relationships of the players in science communication". It has now become possible for all researchers to use the Web to help promote their research. There is a strong common sense argument for using the Web. Publication is free to academics, at least in the richer countries, and so the main cost would frequently be in the design and production of the promotional material. As discussed below, Web publication gives potential access to new audiences. Moreover, it is fast compared to most print media and admits hypertext-specific devices such as linking to full journal or conference papers from publication lists or summaries, copyright permitting. An additional argument for Web publication is the relative ease with which its online impact can be assessed.

Much recent research has investigated the kinds of information about scholarly activities that can be extracted from the Web, particularly Web links (Rousseau, 1997; Ingwersen, 1998; Thelwall, 2000, 2001a-b, 2002a-b). It has been shown in several national university systems that counts of links between universities can produce results that correlate significantly with source and target institutional research productivity (Thelwall, 2001b, 2002a-b; Smith and Thelwall, 2002), which gives some evidence that link counts may be meaningful indicators of scholarly impact. An exercise that attempted to attribute reasons for such link creation found that almost 90% were created for reasons that were associated with scholarly activity but were not online equivalents of bibliographic citations (Wilkinson *et al*., 2002). As a result, Web links can be used to provide evidence of some aspects of informal online scholarly impact. Another potential source of impact information is the Web server log file, which is potentially also useful but does not appear to have been evaluated in a

[1] Thelwall, M. (2002). Research dissemination and invocation on the Web, *Online Information Review*, 26(6), 413-420.

specifically academic context. This can have two purposes: to monitor the effectiveness of dissemination strategies; and to help report on the impact of research to funding bodies. In this paper the potential benefits of online research dissemination will be discussed along with techniques for reporting and assessing the online impact of research. The first issue addressed concerns the potential audiences for academic information.

Who Looks for Research Online?

Other Scholars Lawrence's (2001) findings that papers that are online are cited more than those that are not from a predominantly computer science corpus presumably indicates that authors find new research through Web searches. One advantage of the Web is instant access in contrast to paper journal articles that could have to be obtained from a library. However, electronic technology continues to evolve, presenting researchers with new opportunities to alter their working practices (Lally, 2001) and so this may change. For example, with the increasing availability of digital libraries from publishers, it may be that authors will tend to search these rather than the general Web because of their higher quality content. The same could also apply to well-organized e-print archives. However, the Web would still be a logical choice if this approach fails, particularly in computer science where there are special online article retrieval tools such as ResearchIndex (http://citeseer.nj.nec.com/). The logic of commercial pricing also means that access to digital libraries may well continue to be only partial with a proportion of universities loosing out, perhaps the poorer ones. Alternative low cost digital delivery mechanisms are being explored (Halliday and Oppenheim, 2001), but these have yet to demonstrate that they are capable of replacing commercial publishers.

Educators Many course reading lists are now online, often containing links to informative Web sites and testifying to educators having searched the Web for useful information for students. Educators are often also scholars and researchers in an alternative role and so are a useful additional audience.

Students This constituency is important as the pool from which future researchers will be drawn. Web access is often freely available in universities and it seems logical to hypothesise that its use by students for study-related purposes is likely to increase over the coming years, although this is likely to be discipline dependent (Kling and McKim, 2000). The many books published about information searching on the Web that are aimed at students (e.g. Cooke, 2001) are a symptom of the importance of this area.

Journalists It is known that science journalists search the Web to find information to help them write articles (Trumbo *et al*., 2001), solving a long standing problem of difficulty of access to information sources on science for this profession (Friedman, 1986). These have the potential provide particularly high profile publicity.

Customers for research expertise The Web could potentially be searched by government departments or businesses in order to find investigators that would be able to solve their problems, resulting in the initialisation of collaborations or the awarding of new contracts for research. Although the author knows of no direct evidence for this actually happening, in the context of the widespread use of the Web it seems to be a logical possibility. If this does happen then it provides an additional incentive to present research achievements online.

The General Public The extent to which scholarly activity interests the general public probably varies by discipline and certainly by topic (Peterson, 2001). A majority of the public appears to have some interest in science, 64% in the USA (National

Science Board, 2000). Much knowledge is too complex to be assimilated by almost anyone outside full-time education, yet there are many popular science books in a variety of areas (Weigold, 2001), testifying to a potential audience for at least some research. The fact that such information is sought online is evident in disciplines like astrophysics where museums and amateur astronomers create Web sites and link to academic information sources (e.g. William Herschel Museum, 2000; Bassetlaw Astronomical Society, 2002). Whilst the public is perhaps not the natural immediate audience for most research, evidence that findings are of wider interest would surely be useful in attracting research support from the public purse and there is arguably also a moral obligation in the case of publicly funded research (Aguillo, 2002). This is recognised by American scientists, 81 percent of whom claim to be willing to make some effort to communicate with the public (Hartz and Chappell, 1997). Legislators in the UK have also shown interest in this subject (Dickson, 2000).

It can be seen that the almost all-encompassing nature of the Web as an information source means that online research information can attract a wide variety of different audiences. An effective dissemination strategy will make the most of this opportunity, including the monitoring and assessing of online impact.

Collecting Data on Web Impact

The two traditional mechanisms for evaluating research are peer review (Anderson, 1991) and citation analysis (Jiménez-Contreras *et al*., 2002), with combinations often being used (Roessner, 2000), although of course research is not always evaluated (Steiner and Sturn, 1995). Both techniques are used at the national level for generic official evaluation exercises and also for individual projects as well as promotion and tenure decisions. Citation analysis is a relatively objective quantitative method but must be implemented carefully (Moed, 2002) and only covers part of the impact of research. For example patent citations are not usually included but could potentially provide evidence of research transfer into industry (Oppenheim, 2000), as could other indicators such as commercial funding agreements. Press cuttings have also been used to assess the public visibility of researchers (Posner, 2001) and against this backdrop the Web presents itself as an additional source of information about more general impact.

Citations can be collected via databases such as the Institute for Scientific Information's citation indexes and press cuttings via databases such as that from LexisNexis (http://www.LexisNexis.com). The Web offers a range of new types of data that are fundamentally different to these. Listed below are widely used techniques that would probably not be considered to be intrusive by Web users. More aggressive techniques do exist (Bennett, 2001), but have ethical issues that make them inappropriate for an academic environment.

Server access logs These show how many times a page has been visited. The results have to be interpreted with caution because spiders may repeatedly visit a site, proxy caches may cause an underestimation of visits and site architecture influences page hits, but can nevertheless with care serve as a useful general guide (Nicholas *et al*., 2002). One common log file analysis tool is NetTracker (http://www.sane.com).

Inlinks A link from another Web site is a potential indicator of esteem (Davenport and Cronin, 2000; Borgman and Furner, 2002). Links can be found by AltaVista (http://www.altavista.com/sites/search/adv?what=web/) or Google (http://www.google.com/advanced_search) advanced searches, for example. Visiting the link source pages can give the context of the link, useful additional information, although this is a labour-intensive process.

Non-link Invocations A non-link invocation is where research has been mentioned online without a link. This may take the form of a standard citation, perhaps in a course reading list, although there are many other kinds (Cronin *et al*., 1998). The ability to track down invocations is totally dependant on word frequency issues since a standard search engine query would need to be used. Authors with unusual last names have a distinct advantage in this, but the same applies to research groups and projects. If simple counts are all that is needed then one technique is to use a standard search engine search but then manually vet a sample of the results to estimate the proportion that are valid (Cronin and Shaw, 2002).

The importance of creating sites that are compatible with search engines is discussed in the appendix, and this is critical for the last two sources of information. Search engine coverage of the Web is far from perfect, however, and this creates unavoidable problems when identifying the *sources* of invocations. Several techniques are available to improve coverage, such as using multiple or meta search engines. Areas outside the Web are also accessible to some search engines, however, such as via the Google Groups search (http://www.google.com/grphp). Some invocations will nevertheless be in areas inaccessible to any search engine, the deep or invisible Web, which nevertheless can contain valuable information (Pedley, 2002). Until initiatives to make the deep Web more accessible come to fruition (Medeiros, 2002), tracking down invocations will require an element of manual selection and searching of databases, perhaps with the help of a specialist tool such as a personal portal (Jascó, 2001).

Constructing a Web Invocation Portfolio

There has been an argument about whether it is worthwhile to collect data from the Web about research impact. Bernie Sloan (2001ab) has promoted the idea that researchers can benefit from collating citation and Web based references to their print and electronic publications. He maintains (impressively long) lists of citations and Web pages that invoke his work. Here is an extract from an email discussion list comment from Tom Wilson (2002), which addresses a similar point.

> My most cited paper is "On user studies and information needs" (1981) – a Web search (using Google) revealed 118 pages that listed the title. The pages were reading lists, free electronic journals, and documents that would never be covered by SSCI [Social Science Citation Index], such as reports from various agencies. SSCI revealed, if I recall aright, 79 citations of the paper. The question is: is the Web revealing impact more effectively than SSCI?

Eugene Garfield (2002), the highly respected founder of the Science Citation Index disagrees.

> the bottom line for the researcher is whether anyone has used his or her basic ideas in ongoing research. Until that day of Nirvana arrives when everything will be searchable on the web I am afraid web searching just won't be an adequate substitute.

The start of this argument does not hold water because the end product of research is not necessarily just more research but can also be other things such as useful products and a better-educated workforce. Moreover, individual scientists may be primarily motivated by investigative curiosity, but peer esteem and public approbation are not only relevant on a personal level but also important in obtaining funding to conduct research and in promotion and tenure decisions (Mulkay, 1979). Although Wilson's book does have a high SSCI citation count of 118, it is very impressive *different* information that this book from many years ago is still on course reading lists. The

answer, however, to Wilson's question above must be "no" for precisely the reason that Garfield gives in his second sentence above: it is not a substitute. The Web does not reveal impact more effectively than the SSCI, rather it is capable of revealing different facets of impact (Cronin and Shaw, 2002). There is a potential overlap between the two, via free online journals indexed in the SSCI, but they are essentially different and complementary. For example the UK clickable map (http://www.scit.wlv.ac.uk/ukinfo/uk.map.html) had had over 6 million hits by 12 June 2002, a clear indication of its high utility. Nevertheless, its design and implementation would probably not constitute research activity but its log file and link based online impact would be an impressive demonstration of the need to sustain it.

The importance of a Web invocation is very context-dependent. For example, a link or mention in an online Nature article would be generally valued more highly than something like a link from a researcher's child's personal home page. It follows from this that a basic reporting technique should be to categorise the invocations in a meaningful way. Here are some examples of potential categories, drawn from experience of visiting link sources as well as Cronin *et al*. (1998). It is recommended that each group or individual should create their own categories to make a natural fit with their results, but future research may produce a recommended standardised set of categories to serve as a default starting point. This would facilitate comparisons.

- Course reading lists
- Formal academic publications, including e-journals and e-conference proceedings
- Informal scientific publications (e.g. Nature Science Update)
- Newsgroup discussions
- Research partners
- Companies
- Governmental bodies
- Interested members of the public
- Topic organised link lists and subject gateways

The categorisation approach is recommended over variants of simple link counts, such as Ingwersen's (1998) Web Impact Factor, although these are suitable for reporting links to very large areas of the Web where it would be impractical to differentiate between link origin types. The above categories are illustrative suggestions that would need to be refined or replaced for any given application.

Although each application would be expected to set its own rules for the presentation and categorisation of information, some illustrative examples will be given. The first one is to illustrate additional information about the impact of traditional journal articles for an individual researcher's CV. For this information need, links and log files would clearly not be appropriate, but Web invocations may still be useful. The results could be used to annotate the relevant place in a CV, perhaps in the format below (hypothetical data).

- Wilson, T. D. (1981). On user studies and information needs. Journal of Documentation, 37(1), 3-15. (*79 ISI citations; 10 free online e-journal citations; present on 25 known current course reading lists; 83 other known mentions on the Web*).

For an online initiative the larger numbers would require a different reporting approach. Below is an example using a hypothetical UK funded gateway site. This needs to be more extensive because of its greater relevance for the evaluation of an

online resource. It would even be possible to add a similar link based analysis to a competing resource for comparison purposes.

Table 1. Breakdown of links to a hypothetical UK funded gateway site.

Source of links	Number
UK Departmental links lists	15
UK Library links lists	60
UK Individual course reading lists	84
UK Online news coverage	34
UK Online academic subject portals and links lists	95
UK Individual academics' bookmarks or links lists	85
UK Commercial courses	5
UK Library links	10
UK School student resources lists	12
UK Others	424
Non-UK	820
Total	1,644

Table 2. Log file analysis for hypothetical UK funded gateway site.

Originating domain	Total unique visitors	Page views per visitor
UK Academic	123,000	8.2
UK Commercial	1,000	3.4
Other UK	6,000	5.7
EDU	45,000	8.3
Others	82,000	7.6
Total from the UK	257,000	7.9

The analysis and interpretation of another source of impact information, Web server log files, has been extensively discussed from an information science perspective (e.g. Nicholas *et al*., 2002). Two important points to note are that: careful analysis is needed to remove spurious entries from log files from crawler visits and multiple access by the same person; limited categorisation is possible, because it will not be able to access even basic information such as country of origin for many visitors. Log files can also be used for more immediate monitoring of the effectiveness of a site, particularly if nobody is visiting it. They can also be used to give a useful insight into the type of information that visitors were searching for if they arrived from a search engine (Thelwall, 2001c).

In summary, a Web Impact Portfolio can contain a mixture of categorised and non-categorised information from Web-links, non-link invocations and log files. It can only be used to show online impact and should be used in conjunction with other impact indicators such as bibliometric citations and perhaps press cuttings. As a final point, clearly common sense is required to assess whether the creation of a WIP is an effective use of time. This will depend on factors such as the size of the body of research, the importance of being able to assess its impact, and the extent to which the researcher or team already possess the skills to disseminate research online and report on its impact.

Validity and Reliability Issues when Comparing Web Invocation Portfolios

When assessing and comparing portfolios, for example in a promotions board or research award decision making group, it is important to interpret Web invocations within an effective conceptual framework so as not to inadvertently disadvantage certain groups.

Most research that is recognised in the time period under consideration as being useful will be invoked in various ways in human communication environments from formal bibliographic citations to casual chats between researchers. Citation analysis can capture the first to some extent but most other 'messages' probably disappear, with the partial exception of Web pages. These are not necessarily more important than other forms of communication but tend to be more permanent and visible. They are, however, only a facet of informal communication and cannot be taken to be what in statistical terminology would be called an 'unbiased estimator' of the relative extent of informal communication. The problem is that some disciplines and research areas use the Web more than others, and this is likely to continue to be the case (Kling and McKim, 2000). Ironically, although those from the hard sciences appear to interlink more on the Web than those from the soft (Tang and Thelwall, 2002), this is the opposite of the pattern for formal references, where the soft tend to cite more (Hyland, 2000) and invoke the authors of their references more directly in the text, conceptualising knowledge more explicitly as socially constructed. Perhaps in the future if Web publishing becomes a standard educational skill then the balance of Web linking will be reversed. Whatever the general trend, however, certain types of research projects will have deliverables that are natural Web link attractors, such as subject gateways, online astronomical databases and online teaching information sites (Thelwall, 2002c). As a result of all these factors, the Web impact of two projects may be totally out of proportion with their total informal impact. This mirrors the differential coverage of science in the press (Weigold, 2001).

As a result of the above discussion it is recommended that when comparing researchers, projects or teams partially through their Web invocations, the extent to which the research has a natural fit to the Web should be taken into account. This is a better option than ignoring the Web as an information source because each invocation is still a tangible piece of evidence of impact.

Afterword

A word of caution should also be given to the future evaluations of Web invocation portfolios: since the Web is an unregulated information source, portfolios should never become the most important factors in a decision. If this were to happen then academics would certainly start to play the game (e.g. Stewart, 2002) and either render the whole exercise meaningless, or necessitate strict guidelines on invocations, including some concerning which types of sources would be considered reliable.

References

Aguillo, I. F. (2002), "Web citation issues", SIGMETRICS archives, http://listserv.utk.edu, 10 July.

Anderson, A. (1991), "No citation analyses please, we're British", *Science*, Vol. 252, pp. 639.

Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A. and Raghavan, S. (2001), "Searching the Web", *ACM Transactions on Internet Technology*, Vol. 1 No. 1, pp. 2-43.

Bassetlaw Astronomical Society (2002). Available: http://astronomy.northnottstoday.co.uk/, accessed: 18 June, 2002.

Bennett, C.J. (2001), "Cookies, web bugs, webcams and cue cats: Patterns of surveillance on the world wide web", *Ethics and Information Technology*, Vol. 3 No. 3, pp. 195-208.

Borgman, C and Furner, J. (2002), "Scholarly communication and bibliometrics", In: Cronin, B. (ed.), *Annual Review of Information Science and Technology 36*, Medford, NJ: Information Today Inc., pp. 3-72.

Brin, S. and Page, L. (1998), "The Anatomy of a large scale hypertextual web search engine", *Computer Networks and ISDN Systems*, Vol. 30 Nos. 1-7, pp. 107-117.

Cooke, A. (2001), *A guide to finding quality information on the Internet: Selection and evaluation strategies*, 2nd Ed. London: Library Association Publishing.

Cronin, B. and Shaw, D. (2002, to appear), "Banking (on) Different Forms of Symbolic Capital", *Journal of the American Society for Information Science*.

Cronin, B., Snyder, H.W., Rosenbaum, H., Martinson, A. and Callahan, E. (1998), "Invoked on the web", *Journal of the American Society for Information Science*, Vol. 49 No. 14, pp. 1319-1328.

Davenport, E. and Cronin, B. (2000), "The citation network as a prototype for representing trust in virtual environments". In: Cronin, B. and Atkins, H. B. (eds.). *The web of knowledge: a festschrift in honor of Eugene Garfield*, Metford, NJ: Information Today Inc. ASIS Monograph Series, pp. 517-534.

Dickson, D. (2000), "Science and its Public: The Need for a 'Third Way'", Social Studies of Science, 30(6), pp. 917-924.

Friedman, S. M. (1986), "The journalist's world". In: Friedman, S. M., Dunwoody, S. and Rogers, C. L. (Eds), *Scientists and journalists: Reporting science as news*, New York: Free Press, pp. 17-41.

Garfield (2002), "Web citation", SIGMETRICS discussion list June 2002, Available: http://listserv.utk.edu.

Garner, J., Horwood, L. and Sullivan, S. (2001), "The place of eprints in scholarly information delivery", *Online Information Review*, Vol. 25 No. 4, pp. 250-256.

Goodrum, A. A., McCain, K. W., Lawrence, S. and Giles, C. L. (2001), "Scholarly publishing in the Internet age: a citation analysis of computer science literature", *Information Processing and Management*, Vol. 37 No. 5, pp. 661-676.

Halliday, L. and Oppenheim, C. (2001), "Developments in digital journals", *Journal of Documentation*, Vol. 57 No. 2, pp. 260-283.

Harnad, S. and Carr, L. (2000), "Integrating, navigating, and analysing open eprint archives through open citation linking (the OpCit project)", *Current Science*, Vol. 79 No. 5, pp. 629-638.

Harter, S. P. and Ford, C. E. (2000), "Web-based analyses of e-journal impact: approaches, problems, and issues", *Journal of the American Society for Information Science*, Vol. 51 No. 13, pp. 1159-1176.

Hartz, J. and Chappell, R. (1997), "Worlds apart: How the distance between science and journalism threatens America's future", Nashville, TN: First Amendment Center.

Hyland, K. (2000), *Discipinary discourses: social interactions in academic writing*, Harlow: Longman.

Ingwersen, P. (1998), "The calculation of Web Impact Factors", *Journal of Documentation*, Vol. 54 No. 2, pp. 236-243.

Jascó, P. (2001), "Savvy searching", *Online Information Review*, Vol 25 No. 4, pp. 267-270.

Jiménez-Contreras, E., López-Cózar, E. D., Ruiz-Pérez, R. and Fernández, V. M. (2002), "Impact-factor rewards affect Spanish research", *Nature*, Vol. 417, pp. 898.

Kleinberg, J. (1999), "Authoritative sources in a hyperlinked environment", *Journal of the ACM*, Vol. 46 No. 5, pp. 604-632.

Kling, R. and McKim, G. (2000), "Not Just a Matter of Time: Field Differences in the Shaping of Electronic Media in Supporting Scientific Communication", *Journal of the American Society for Information Science*, Vol. 51 No. 14, pp. 1306-1320.

Lally, E. (2001), "A researcher's perspective on electronic scholarly communication", *Online Information Review*, Vol. 25 No. 2, pp. 80-87.

Lawrence, S. L. (2001), "Online or invisible?", *Nature*, Vol. 411, pp. 521.

Medeiros N. (2002), "Reap what you sow: harvesting the deep Web", *OCLC Systems & Services*, Vol. 18 No. 1, pp. 18-20.

Moed, H.F. (2002), "The impact-factors debate: the ISI's uses and limits", *Nature*, Vol. 415, pp. 731-732.

Mulkay, M. (1979), *Science and the sociology of knowledge*, London: Allen & Unwin.

National Science Board (2000), "Science and Engineering indicators 2000", Available: http://www.nsf.gov/sbe/srs/seind00/, accessed 2 July, 2002.

Nicholas, D., Huntington, P. and Williams, P. (2002), "Evaluating metrics for comparing the use of Web sites: a case study of two consumer health web sites", *Journal of Information Science*, Vol. 28 No. 1, pp. 63-75.

Oppenheim, C. (2000), "Do patent citations count?". In: Cronin, B. and Atkins, H. B. (eds.). *The web of knowledge: a festschrift in honor of Eugene Garfield*. Metford, NJ: Information Today Inc. ASIS Monograph Series, pp. 405-432.

Pedley, P. (2002), "Why you can't afford to ignore the invisible Web", *Business Information Review*, Vol. 19 No. 1, pp. 23-31.

Peterson, I. (2001), "Touring the Scientific Web", *Science Communication*, Vol. 22 No. 3, pp. 246-255.

Posner, R. A. (2001), *Public intellectuals: A study of decline*, Cambridge, MA: Harvard University Press.

Pringle, G., Allison, L. and Dowe, D. L. (1998), "What is a tall poppy among web pages?" *Computer Networks and ISDN Systems*, Vol. 30 Nos. 1-7, pp. 369-377.

Roessner, D. (2000), "Quantitative and qualitative methods and measures in the evaluation of research", *Research Evaluation*, Vol. 8 No. 2, pp. 125-132.

Rousseau, R. (1997), "Sitations, an exploratory study", *Cybermetrics*, Vol 1. Available: http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html

Sloan, B. (2001a), "Personal Citation Index: Exploring the Impact of Selected Papers", Available: http://www.lis.uiuc.edu/~b-sloan/pci2.html. Accessed 18 June, 2002.

Sloan, B. (2001b), "Personal Citation Index", JESSE archives November 2001 (#74). Available: http://listserv.utk.edu.

Smith, A. and Thelwall, M. (2002), "Web Impact Factors for Australasian Universities", *Scientometrics*, Vol. 54 Nos. 1-2, pp. 363-380.

Steiner, M. and Sturn, D. (1995), "Elements of evaluation of science and technology policy in Austria", *Research Evaluation*, Vol. 5 No. 1, pp. 98-108.

Stewart, I. (2002), "Reassessing Research Assessment in the UK", *Science*, Vol. 296 No. 5574, pp. 1802.

Sullivan, D. (2001), "Search Engine Placement Tips", *Search Engine Watch*. Available: http://www.searchenginewatch.com/webmasters/tips.html, accessed 24 July, 2002.

Thelwall, M. (2000), "Web Impact Factors and Search Engine Coverage", *Journal of Documentation*, Vol. 56 No. 2, pp. 185-189.

Thelwall, M. (2001a), "Results from a Web Impact Factor crawler", *Journal of Documentation,* Vol. 57 No. 2, pp. 177-191.

Thelwall, M. (2001b), "Extracting Macroscopic Information from Web Links", *Journal of the American Society for Information Science and Technology*, Vol. 52 No. 13, pp. 1157-1168.

Thelwall, M. (2001c), "Web Log File Analysis: Backlinks and Queries", *ASLIB Proceedings*, Vol. 53 No. 6, pp. 217-223.

Thelwall, M. (2002a), "A comparison of sources of Links for academic Web Impact Factor calculations", *Journal of Documentation*, Vol. 58 No. 1, 60-72.

Thelwall, M. (2002b), "Conceptualizing documentation on the Web: an evaluation of different heuristic-based models for counting links between university web sites", *Journal of the American Society for Information Science and Technology*, 53(12), 995-1005.

Thelwall, M. (2002c), "Subject gateway sites and search engine ranking", *Online Information Review*, Vol. 26 No. 2, pp. 124-138.

Town, W. G., Vickery, B. A., Kuras, J. and Weekes, J. R. (2002), "Chemical e-journals, chemical e-prints", *Online Information Review*, Vol. 26 No. 3, pp. 164-171.

Trumbo, C. W., Sprecker, K. J.; Dumlao, R. J., Yun, G.W. and Duke, S. (2001), "Use of E-Mail and the Web by Science Writers", *Science Communication*, Vol. 22 No. 4, pp. 347-378.

Weigold, M. F. (2001), "Communicating science: A review of the literature", *Science Communication*, Vol. 23 No. 2, pp. 164-193.

Wilkinson, D., Harries, G., Thelwall, M. and Price, E. (2002), "Causes of Academic Web Site Interlinking: Evidence for the Web as a Novel Source of Information on Informal Scholarly Communication", University of Wolverhampton.

William Herschel Museum (2002). Available: http://www.bath-preservation-trust.org.uk/museums/herschel/index.html, accessed: 18 June 2002.

Wilson, T. (2002), "Web citation", JESSE discussion list. Available: listserv.utk.edu

Ziman, J. (1992), "Not knowing, needing to know, and wanting to know". In: B. V. Lewenstein (Ed.), *When science meets the public*, Washington, DC: American Association for the Advancement of Science, pp. 13-20.

Appendix: Using the Web to Publicise Research

There are many different ways in which research can be presented on the Web and it seems that publicly accessible academic papers on the Web attract relatively little online attention, at least in terms of inlink counts (Thelwall, 2001b). Information about research, such as lists of publications and general descriptions, can often be

found in departmental pages, research group pages and individual scholars' pages. The key issue in any context is how a new user will find the information. Typically this may well be through the major Web information retrieval tool; the search engine. As a result, steps should be taken to ensure that the content is found and "understood" by search engines so that it is returned by appropriate searches. This means using indexable HTML (not dynamic HTML, PDF or shockwave, for example) and using appropriate HTML tags including title and headings to effectively describe contents (Pringle *et al*., 1998). Additionally, the use of Frames based sites should be avoided because these can cause complications for search engine indexes that can result in the pages being ignored (Sullivan, 2001). Similarly, there are other forms of Web publishing that can cause problems, including the storing of pages in databases that will not be indexed by search engines. In the era of link based ranking algorithms, simply making sure that pages are available to search engines and adequately described for them is not enough, however. A page or collection of pages should also signal its content type by linking to similar pages elsewhere on the Web and, if possible, attracting reciprocal links (Brin and Page, 1998; Kleinberg, 1999; Arasu *et al*., 2001).

One generic problem with search engines is that they are susceptible to word ambiguity and liable to poor results for frequently occurring terms. For example, a search for "statistics" will return millions of pages of server statistics. A consequence of this is that word frequency can be an issue in certain circumstances. A case in point is research project titles and acronyms. A well known project will be difficult to find on the Web when its name is a common term, unless its URL is known. There is a case, then, for choosing deliberately obscure names to facilitate discovery through online searching, if the name of a project is expected to be widely known. This has the additional benefit that it will be easier for the project participants to identify web invocations of their work, especially when these are unaccompanied by links to the home site.

The general rule for web publicity must be that more is better. All researchers and groups can have Web pages giving full bibliographic references for their work. In some cases publishers allow articles to be posted on the Web in personal web sites, giving instant full access to visitors. General descriptions of research may also by helpful for journalists, potential clients and PhD students to assess the researcher(s) or their work. Science communication research has suggested that a good perspective on which to base marketing strategies is attempting to answer in concrete terms the question, "What do people want to know in their particular circumstances?" (Ziman, 1992).