University of Portsmouth

Research Issue 2009

Harvest

UNIVERSITY LIBRARY NEWSLETTER

MAKING OUR RESEARCH MORE VISIBLE: DEVELOPING OUR INSTITUTIONAL REPOSITORY

Institutional Repositories (IRs) collect, organise, describe, store, preserve and make accessible an institution's scholarly output in digital format. Making research openly available on the web in this way leads to greater exposure to search engines such as Google, which in turn leads to higher citation rates[1]. This is of importance for the next Research Excellence Framework (REF), which will be based partly on bibliometric indicators, i.e. the number of times an article is cited[2].

Portsmouth's experience so far has been limited. We have had a small repository since 2004 (eprints.libr.port.ac.uk), but the REF is now driving its relaunch and revitalisation to collect our output in one accessible place and improve citation rates. Recent development of international standards allows a number of IRs to be searched simultaneously and text-mining offers the possibility of finding new correspondences between areas of research (www.nactem.ac.uk). Therefore, a centralised collection makes our output more searchable and structured than uploading to individual departmental web pages.

IRs can include a variety of documents including journal articles, books, grants information, conference proceedings and theses, as well as conference presentations, descriptions of artefacts, film, music and other media. It is for the individual institution to define the types of format to be included. In the case of a journal article, the standard unit of research output in many disciplines, the IR entry would describe the article, include assigned headings for standard retrieval across an IR or a group of IRs (such as whether the journal uses peer review) and offer a link to the full text of the article. Many universities require staff to deposit their research outputs in the IR either in the form of pre-prints (the final version of the article before it goes to the journal for publication, i.e. after peer review) or as post-prints (once the copyright restrictions of publishers have been met, e.g. some require a six-month embargo post-publication).

Open access

Allied to the development of IRs is the open access movement. This started in response to the situation where academics write for journals then give away the copyright, so libraries then have to buy back the content. A number of universities have therefore adopted an open access policy, which encourages academics to make their research available freely through their IR. There are different models, including one where the author pays up front for publication (encouraged by some grant-making bodies). Authors are also encouraged to negotiate with journals for a less restrictive copyright agreement. Open access has advantages for countries in the developing world which can't afford expensive academic journals (and this may have wider applicability in these recessionary times).

A group led by David Arrell is looking at developing a new IR for Portsmouth. We will be in touch over the next few months about how you can contribute, but if you have questions, ideas or opinions please contact me.

Roisin Gwyer
University Librarian
Ext.3222
roisin.gwyer@port.ac.uk

References

1 Norris, M., Oppenheim, C. and Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology*, 59 (12): 1963–72

2 www.hefce.ac.uk/Research/ref

Related websites

University of Southampton IR: www.eprints.soton.ac.uk

Publishers' copyright permissions: www.sherpa.ac.uk/romeo.php

Funding bodies' policies: www.sherpa.ac.uk/juliet/index.php

List of open access repositories: roar.eprints.org

List of open access journals: www.doaj.org

CITATION SEARCHING: A USEFUL RESEARCH TOOL

When structured searching for journal articles via keywords or subject headings fails, citation searching can often locate useful articles. It can also reveal how important a particular article might be.

What is citation searching?

When you have identified a key article, it is common to look at the references the author has used and maybe read one or two of the key documents. This gives you an insight into the historical development of the subject you are exploring.

However, with some databases it is possible, by inverting the above process, to see who has listed the article in their bibliography. This has the advantage of moving you forwards in time and possibly finding totally new areas of research that you may not have considered. Locating new references by finding an earlier reference that they cite is referred to as bibliographic coupling.

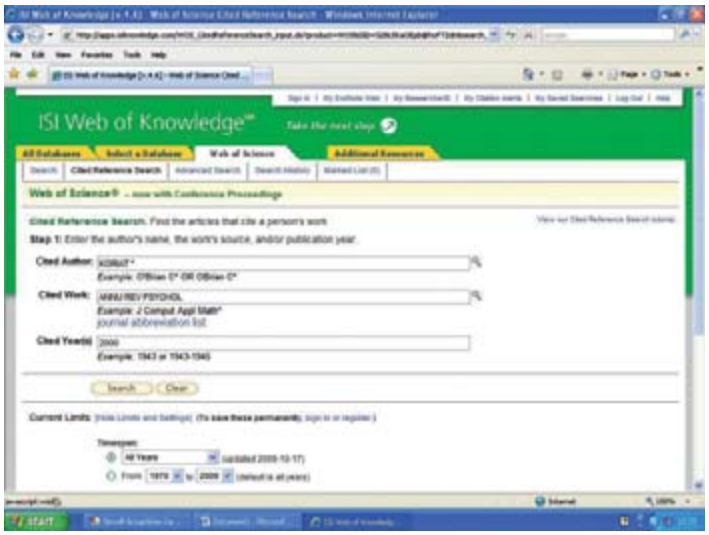
For example, the following article has been found to be key for those doing research on memory:

Toward a psychology of memory accuracy
Koriat A, Goldsmith M, Pansky A
Annual Review of Psychology
Volume: 51
Pages: 481–537
Published: 2000

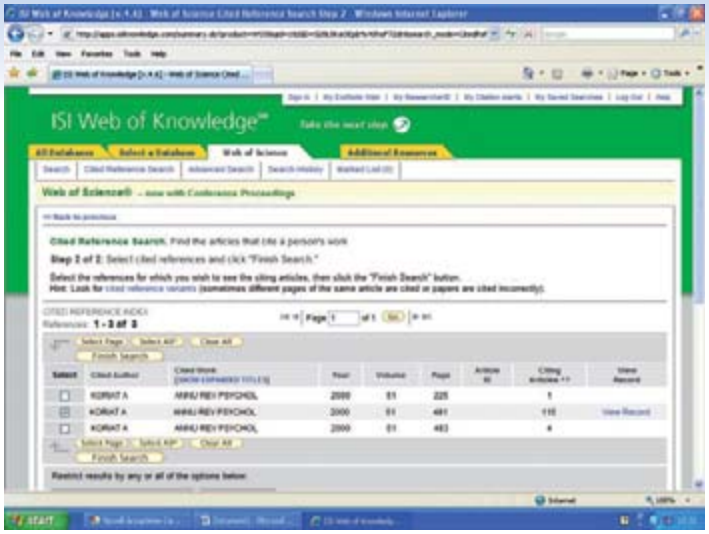
In this article's bibliography the authors list almost 14 pages of references with three articles published in 2000. The rest of the citations stretch back to 1950 and earlier. Which is fine for a start, but even more interesting is to do a search for other articles where this work has been cited.

How does one search?

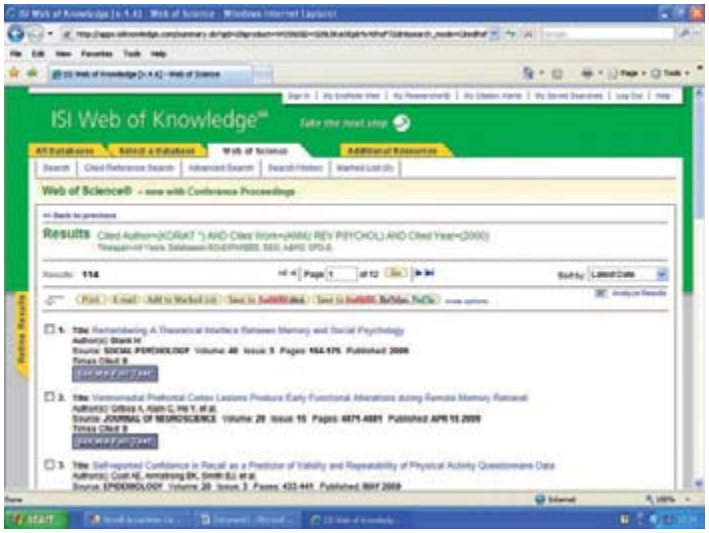
One of the largest databases supplying a cited reference search option for articles dating back to 1971 is Web of Science, which is part of Web of Knowledge (wok.mimas.ac.uk). The search using the above example is illustrated below:



Follow the pattern (shown above) for data entry carefully. This search resulted in three results as shown above right:



Once identified, select the article and then finish the search.



The search lists 113 articles published from 2000 to the present day that have quoted the above article, including eight articles published in 2009.

Other databases

Other bibliographic databases offer the facility to search cited references, although this option may not be as clear or well developed as it is in the Web of Knowledge. For example, in Cambridge Scientific Abstracts databases the option to search references appears, but only if you select a database that offers such a feature. PsycInfo implements this feature, Medline does not.

Sue Stevenson
Assistant Faculty Librarian (Science and Technology)
Ext.3236
sue.stevenson@port.ac.uk

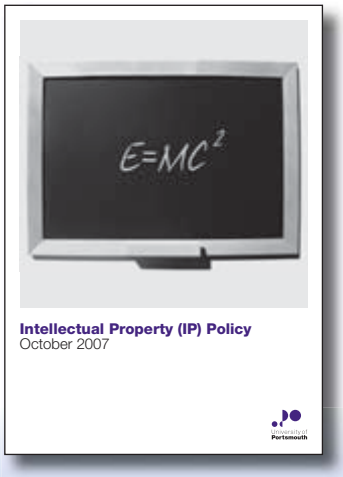
COPYRIGHT: NOT SEXY, BUT NECESSARY

UK copyright is governed by the 1988 Copyright, Design and Patents Act (CDPA). Copyright protection is automatic and, in the UK, there is no registration or other procedure to follow; protection automatically exists as soon as an original work is recorded in some material form. As a property right, copyright can be exploited, bought, sold, bequeathed, rented or licensed, just like any other property. Researchers may have copyright concerns in two areas: their own copyright and their ability to use the copyrighted material of others. There follows a brief look at some of the issues, with pointers to further information.

Do I own the copyright to my own work?

The initial owner of the copyright in a literary, dramatic, musical or artistic work is the author or creator, unless the work was done in the course of employment. In this case, the employer owns the copyright, unless a contract specifies otherwise.

- The document *Intellectual Property (IP) Policy*, October 2007, lays out the University's policy on the ownership, protection and exploitation of the intellectual property produced by staff of the University.
- Thody and Thody (2006, pp.227–230) have a useful FAQ section on the use you can make of your thesis and on granting other people permission to use your work.
- Copyright can be assigned to another party. For example, in many cases the author of a published journal article does not hold the copyright, but has assigned it to the publisher. As rights agreements vary from publisher to publisher, you'll need to check if you are submitting work for publication.



Can I use other people's work to inform my research?

Copying other people's work for purposes of 'research or private study' is generally taken to be limited to the following:

- One article from any one issue of a journal (even if that one article is the whole issue).
- One chapter or up to five per cent (whichever is the greater) of a book or similar publication.
- One poem or short story of up to ten pages from an anthology or the report of one case in law reports.

The University of Portsmouth has signed up to a number of licences which enable you to copy for educational/research purposes. These are listed on our copyright web pages at **www.port.ac.uk/library/copyright**.

If you need to copy outside the guidelines provided by the licencing body or find that a publisher is not covered by any of the licences the University is signed up to, then you will need to seek permission to copy from the publisher or copyright holder directly. This can take some time, particularly if you aren't sure who the copyright holder is.

Just as an aside, works on the internet are not freely available to copy or use. Instead, you'll need to check the home site for links such as those to 'Copyright', 'About this site', 'Info' or 'Contact webmaster' and, if there is no statement giving permission for the use you want to make of the page, you will have to obtain permission.

Can I use other people's work in my work?

You can quote or include short extracts or ideas in your work which should be acknowledged in your references.

It's usually clear on the item if the rights holder does not allow you to use their material, but if you are unsure, you would be best advised to contact them to obtain permission to use their work. Thody and Thody (2006, p.225) supply a helpful checklist for requesting the use of copyrighted material and comment that many copyright owners are pleased to have their material included in another work. If you do however initially receive a negative response, they do recommend trying persuasion by offering some form of compensation, but ultimately warn that 'if you cannot locate or persuade a copyright owner, do not include the material. Remember: if in doubt, leave it out!'

COPYRIGHT: NOT SEXY, BUT NECESSARY (CONTINUED)

One additional thing to consider is from Rumsey (2008, p.146), which warns researchers to remember that 'although material might be permitted to be included in a thesis for the purposes of examination, it does not automatically imply that the material may be made freely available via the internet by the author of the thesis.' You might therefore consider also asking permission from the rights holder to include their work in any future online version of your thesis.

At this point, plagiarism – the passing off of someone else's work or ideas as your own or the use of someone else's work without crediting that person – rears its ugly head. Rumsey (2008, p.152) believes that 'most researchers do not need reminding that it is good practice and courteous to acknowledge the work of others', but goes on to supply a useful exploration of the issue. Thody and Thody (2006, p.226) also warn the researcher that, whilst it would not be plagiarism to use your own material from another source (such as a thesis or article), that in doing so, there may instead be copyright issues if you are not the copyright owner of the first source.

References and further information

Rumsey, S. (2008). *How to find information: a guide for researchers*. (2nd Ed.) Maidenhead: McGraw-Hill: Open University Press.

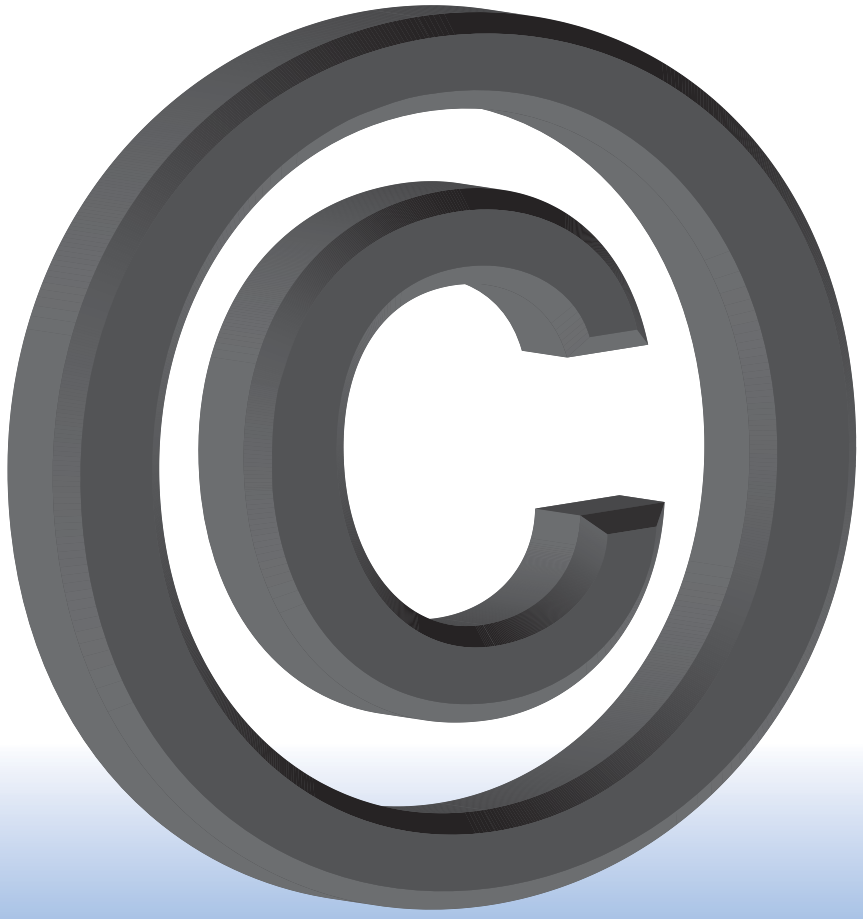
Thody, L. S. and Thody, S. (2006). Standing on the shoulders of giants – without violating their copyright. In: A. Thody *Writing and presenting research*. (pp.221–231) [Electronic version] London: Sage Publications.

University of Portsmouth. *Copyright Code*, September 2007
www.port.ac.uk/accesstoinformation and select 'C' in the Document Warehouse A–Z list and from there 'Copyright Code'.

University of Portsmouth. Intellectual Property (IP) Policy, October 2007
www.port.ac.uk/accesstoinformation and select 'I' in the Document Warehouse A–Z list and from there 'Intellectual Property (IP) Policy'.

University of Portsmouth Library
www.port.ac.uk/library/copyright

Kath Shakespeare
Loans Librarian
Ext.3245
kath.shakespeare@port.ac.uk



WEB 2.0: COOK, HOROLOGIST, PLUMBER OR MURDERER?

Over the past three months colleagues in the University Library have commented that my conversation has been dominated by four topics: mashups, ticTOCs, pipes and The Body in the Library. I must admit that seeing a list like this makes me realise how eccentric I must have seemed. However, I thought that they deserved a fuller explanation, and that others might be interested in what I've learned in a relatively short space of time.

RSS (Really Simple Syndication)

I should say that whilst I was aware of RSS feeds, I had always had a lingering feeling that they were related to repetitive strain and certainly not relevant to my work. However, having become involved with the development of Subject Hubs for the Institute for Criminal Justice Studies (ICJS) students in Victory, I realised that staff saw RSS feeds as a way of keeping the sites up-to-the-minute and providing an added reason for repeated visits. Feeds are particularly useful as contributions to ICJS's knowledge transfer projects.

Working together with colleagues in ICJS, I developed three brands of tailored feeds for specific Subject Hubs: current, academic and official news. My ICJS colleagues came up with an incredibly challenging topic for me to start with: diversity. I selected ten general news sources and identified relevant official bodies and harvested feeds from the top twenty list of journals ICJS provided. I then set to work identifying suitable terms to use as filters for the feeds. After a lot of experimentation I finally created a usable filter for 'diversity' using 69 key terms (for other subjects a dozen or so have proved effective). This was probably the most difficult part of the whole process.

Mashup

Groundwork in place, I then ventured into the Web 2.0 realm of the mashup. This is the combining of applications or information (in this case RSS feeds) from a variety of sources to create a new service.

First I located all of the feeds from my various sources. This is easy enough to do for news sites and government departments, but they are often disparate in coverage, frequency and presentation, so my final filters needed a little tweaking. Academic journals present more of a challenge, so I was pleased to discover an easier way to source the majority of them. I had heard other librarians talking about ticTOCs at a Joint Information Systems Committee (JISC) consultation meeting. This site provides easy access in one place to RSS feeds from 12,387 ejournals harvested from publishers, complete with the abstract where available. These are searchable by title, publisher or subject and you can view them in situ.

I chose to export the feeds to manipulate them. To do this I linked the feeds into an application called Yahoo! Pipes™ and 'piped' them through my chosen filters, plus some additional filters, to sort and weed out duplicate entries. I was then able to provide ICJS with a new amalgamated feed for their chosen topic.

So, while I still can't cook, mend a clock or change a washer, I am, thanks to my skills with old-fashioned Boolean logic (plus some help from some really useful techie tools), able to produce effective subject-orientated news feeds. What you may be left wondering about is that Body in the Library. Keep an eye on the Library website and Victory for more news on this in the near future.

Linda Jones
Law Librarian
Ext.3240
linda.jones@port.ac.uk

JOURNAL IMPACT FACTORS: MEASURE OF QUALITY

Viewed as measures of journal quality, impact factors can be useful tools for the management of a library's journal collections. They show the average number of times that articles in a journal are referenced by other articles.

Journal impact factors were developed by Eugene Garfield in the 1960s as an aid to evaluating journals for inclusion in his publications *Current Contents*™ and *Science Citation Index*™ (Garfield, 1998, 2006). Garfield wanted an average citation measure that would account for the effects of the size and age of a journal on the total number of citations it receives. For example, older, larger journals generally receive more citations because they have larger bodies of articles available for citing.

Today, the term 'impact factor' usually refers to figures calculated and published by Thomson Reuters each year in the Science and Social Sciences Editions of *Journal Citation Reports (JCR)*™.

Calculation

Calculating an impact factor requires a **denominator** (total number of articles published) and a **numerator** (total number of citations to those articles) with time periods being defined for both variables. The **publication window** is the period during which the articles included in the calculation were published (usually the two previous years); the **citation window** is the period during which citations to these articles were counted (usually the journal impact factor year).

As an example, the 2007 impact factor (released 2008) for *Advances in Physics* was calculated as follows:

- Number of 2007 citations to articles published in 2005–2006 in *Advances in Physics* = 201
- Number of articles published in 2005–2006 in *Advances in Physics* = 21
- 2007 impact factor for *Advances in Physics* = 201/21 = 9.571

Issues

There are several issues associated with the use of impact factors as measures of quality, and it is important to understand these before utilising impact factor data in any decision-making process. Some are outlined below:

Subject variation in journal impact factors

The average number of citations to articles during the two years after publication varies across subject fields. These differences in citation activity lead to impact factors of different magnitudes between different disciplines. For example, the average journal impact factor for all the cell biology journals covered by *Journal Citation Reports (JCR)*™ is around 5.8; the same figure for history journals is around 0.4; different subjects, different impact factors. Some possible reasons for different citation levels are as follows:

- **Variation in the average number of authors per article.** Highly-cited articles tend to have more authors (Aksnes, 2003) and subject fields with more authors per paper tend to have higher impact factors (Amin and Mabe, 2000).

- **Publication behaviour in different subject fields.** Social scientists make greater use of books as a method of disseminating information than do the natural sciences. This means that a relatively high proportion of references from journal articles in these fields are to books (Garfield, 1976), reducing the total number of inter-journal citations. So it is too in engineering, with many citations going to conference proceedings rather than to journals (Harzing and van der Wal, 2008). Basic research fields also tend to receive more citations than related applied fields and, therefore, have higher impact factors. For example, journals in basic medical fields generally have higher impact factors than those in clinical medical fields (Seglen, 1997).

- **Citation data coming only from publications that are indexed by Thomson Reuters in the Journal Citation Reports.** Differences in the level of coverage by Thomson Reuters across subject fields are another cause of the varying magnitudes of citation activity as measured by *Journal Citation Reports (JCR)*™ (Althouse, 2008).

JOURNAL IMPACT FACTORS: MEASURE OF QUALITY (CONTINUED)

In addition to different levels of recorded citation activity between fields, different citation patterns also affect the magnitude of impact factors. Only citations received in the two calendar years after publication count towards a journal's impact factor, and the proportion of citations falling within this impact factor window varies considerably between subject areas. Approximately 22 per cent of citations to cell biology articles fall within this window compared to eight per cent of citations to articles in economics or mathematics. This compounds the difference in impact factors between these fields since cell biology articles not only receive more citations in total, but a higher proportion of these contribute towards impact factors (Althouse, 2008).

Why two-year impact factors?

Journal impact factors were originally devised to help with journal selection for inclusion in *Current Contents*™. At the time of the inception of impact factors, the primary fields of focus for *Current Contents*™ were molecular biology and biochemistry (Garfield, 1998). In these fields, 25 per cent of citations received in a particular year were accounted for by articles published in that year and the two previous years (Garfield, 2003). Because of this, using a measure that only included citations to recent articles was considered appropriate.

Thomson Reuters justifies its continued publication of two-year impact factors on the basis that it is both current and responsive (Althouse, 2008). Using citations in the current year (where 'current year' refers to the impact factor year rather than the year of *Journal Citation Reports (JCR)*™ publication, which is the following year) and articles in the two previous years, ensures that *Journal Citation Reports (JCR)*™ impact factors are sensitive to recent changes in citation activity. For example, a journal's impact factor will respond to papers or topical issues that receive an abnormally high number of citations, but this effect will last for no more than two years. A longer publication window would dilute these effects, but would include them for a longer period.

Thomson Reuters considers this sensitivity to be one of the strengths of the two-year impact factor. However, it could also be considered a weakness since it means that several years of impact factors need to be considered to gauge the general impact of a title. Another advantage of the *Journal Citation Reports (JCR)*™ impact factor is that it only requires three years' worth of data (two publication years and one citation year) to calculate.

Five-year or seven-year impact factors have been suggested, but these would take over twice as long to produce and this delay could be detrimental to new journals trying to establish themselves, and currently the two-year measure is the only official impact factor published in the annual *Journal Citation Reports (JCR)*™.

Consequences of subject variations

The variation in average impact across different subjects means that impact factors cannot be used to compare journals from different subject areas. *Journal Citation Reports (JCR)*™ is classified into subject categories, but it is only at a broad level that journals are ranked according to their impact factors (Pringle, 2008). Even within these subject categories some journals will have a subject advantage over others. For example, this can apply where a subject category contains both basic research and applied journals. There can also be a disadvantage in categories which are multidisciplinary. All these factors need to be considered when comparing journal rankings within a subject category.

Subject variation in impact factors means that it is unwise to create an 'average impact factor' for a publisher's entire list of journals. Now that many publishers offer bundled sales deals, it can be tempting to try and gauge the quality of such an offering by creating an average impact factor for the journals included in the deal. However, if subject differences mean that the impact factors of journals from different fields cannot be compared, then the data should not be combined either. Publisher-wide averaging of impact factors will always favour those publishers with strong life sciences programmes over those with strong social science programmes. Averaging impact factors across even a single subject package is unlikely to give a fair view of the quality of the package. The same is true of trying to compare publishers on the basis of price per impact factor where journals from multiple *Journal Citation Reports (JCR)*™ categories are included in the analysis, for example the LISU report *Trends in Scholarly Journal Publishing*, which combines diverse subject categories into a single analysis on biomedical titles (White and Creaser, 2007).

JOURNAL IMPACT FACTORS: MEASURE OF QUALITY (CONTINUED)

Is every published item counted?

Many journals publish rarely-cited ('non-source') items such as news articles, editorials, letters to the editor and meeting abstracts. So that journals publishing this material are not unduly penalised, they are not counted in the article total for impact factor calculations. However, although these items are rarely cited, there are exceptions and these citations do count towards the citation total.

The consequence of this discrepancy is that journals publishing a large number of non-source items or journals publishing particularly interesting 'non-source items', can have artificially inflated impact factors. For these journals, citations are being counted in the numerator to articles that are not counted in the denominator. The number of non-source items published by particular journals can be found under the heading 'Journal Source Data' in *Journal Citation Reports (JCR)*™. The heading here 'Other Items' refers to the number of non-source items, although there is no indication of the number of citations these articles have received.

Review articles

Authors will often cite one review article rather than the many primary research articles it is based on, so review articles are generally cited more often than primary research articles. Therefore, review journals usually have higher impact factors than other journals in their field. For example, the aggregate 2007 impact factor for science journals publishing over 75 per cent of review articles is 2.5 times higher than the figure for the remaining journals. *Journal Citation Reports (JCR)*™ does not list review journals and primary journals separately, therefore review journals and those containing a high proportion of review content are often ranked amongst the highest journals in their fields. For instance, the top three journals in the 2007 toxicology category are review journals and the fourth-ranked journal publishes over 25 per cent of review articles. So, it is important to consider the proportion of review content when comparing the impact factors of different journals.

Skewed citation distribution

Impact factors are designed to be 'a measure of the frequency with which the 'average article' in a journal has been cited in a particular year' (White and Creaser, 2007). However, the distribution of citations to articles within a journal is generally highly skewed with a minority of articles receiving the majority of citations, especially within the narrow impact factor windows (Seglen, 1997). What this means is that the impact factor often does not actually give a good indication of the frequency with which the 'average article' has been cited. There have been calls for the production of a median-based average for journals along with the mean (Rossner, Van Epps and Hill, 2007), but this would not be possible without a change in the way the data is collected for *Journal Citation Reports (JCR)*™. Data is currently collected at the journal level and therefore measures which rely on distribution data such as medians cannot be calculated from the current data set.

Journal size and impact factor variability

Impact factors vary from year to year, so it is important not to read too much into subtle changes in these figures, especially in smaller journals. Impact factors can be thought of as the mean number of citations to a biased sample of articles from the population of all articles in that field (Amin and Mabe, 2000). Statistically, smaller samples will have greater sampling errors than larger ones, the mean values delivered by repeated samplings being more variable. A small change in the total number of citations to a journal that publishes

few articles a year will lead to a large change in the impact factor between years, especially true if the journal has a low impact factor to start with. It only takes two extra citations to a journal with an impact factor of 0.500 which publishes 20 articles a year (and so would have an impact factor denominator of 40) to increase its score to 0.550, a ten per cent increase. A larger journal, one publishing 100 articles a year with an impact factor of 4.000, would need to gain an extra 80 citations to see a ten per cent increase. It is important to bear this caveat in mind when evaluating any changes in impact factor between two years. If a journal is very small, then a relatively large change in impact factor (or a consistent increase or decrease over several years) is required to demonstrate that there has been an underlying change in the population from which the journal draws its articles.

In addition to averaging impact factors across a list, it might be tempting to calculate the mean percentage increase of a group of titles in an attempt to gauge the improvement of the list as a whole. However, the high variability in impact factor is one reason to avoid this. In the examples given above, both journals see an increase of ten per cent for very different changes in the actual number of citations received between the two years. Both these changes would be given exactly the same weight in a simple average of the percentage changes in impact factors across a list. There is also a definite increase bias in such a calculation. Suppose that two journals both publish 20 articles per year, giving both titles an impact factor denominator of 40 (2 x 20), and that one gains six citations and the other loses six. The change in average impact factor is zero, but the mean percentage change is +45 per cent because the increase in citations leads to a 150 per cent increase in impact factor compared to only a 65 per cent decrease for the loss of citations.

What this all means is that simple mean percentage increases across a list can be highly skewed by small journals which have seen a high percentage change in impact factor due to only a small, and quite possibly random, change in the number of citations received. This skew is even greater when the change is an increase.

Further information

This article is not intended to be a comprehensive look at the many issues that surround the many different ways in which this ubiquitous measure is used. Those interested in such issues may wish to follow some of the references which are listed below.

Andrew Barrow
Faculty Librarian (Science)
Ext.3236
andy.barrow@port.ac.uk

References

Aksnes, D. (2003). Characteristics of highly cited papers. *Research Evaluation*, 12(3), 159–170.

Althouse, B. (2008). Differences in impact factor across fields and over time (Department of Economics, UCSB, Paper 2008-4-23). Santa Barbara, CA: University of California at Santa Barbara.

Amin, M. and Mabe, M. (2000). Impact factors: use and abuse. *Perspectives in Publishing 1* Retrieved 16 June 2009, from http://www.elsevier.com/framework_editors/pdfs/Perspectives1.pdf

Garfield, E. (1976). On the literature of the social sciences and the usage and effectiveness of the Social Sciences Citation Index. *Current Contents*, 34 (August 23rd), 5–10.

Garfield, E. (1998). The use of journal impact factors and citation analysis for evaluation of science. *41st Annual Meeting of the Council of Biology Editors* Retrieved 23 June 2009, from http://www.garfield.library.upenn.edu/papers/eval_of_science_CBE(Utah).html

Garfield, E. (2003). The meaning of the impact factor. *International Journal of Clinical and Health Psychology*, 3(2), 363–369.

Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA (Journal of the American Medical Association)*, 295(1), 90–93.

Harzing, A.-W. and van der Wal, R. (2008). Google Scholar as a new source for citation analysis? *Ethics in Science and Environmental Politics*, 8(1), 62–71.

Pringle, J. (2008). Trends in the use of ISI citation databases for evaluation. *Learned Publishing*, 21, 85–91.

Rossner, M., Van Epps, H. and Hill, E. (2007). Show me the data. *Journal of Cell Biology*, 179(6), 1091–1092.

Seglen, P. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ (British Medical Journal)*, 314, 497.

White, S. and Creaser, C. (2007). *Trends in Scholarly Journal Prices 2000–2006* (LISU Occasional Paper No.37). Retrieved 23 June 2009, from www.lboro.ac.uk/departments/ls/lisu/pages/publications/oup2.html

IMPACT FACTORS: THE H-INDEX

The h-index, developed by Jorge Hirsch, attempts to measure the impact of an individual researcher based upon a set of their most cited papers and the number of citations they have received. Although most commonly calculated for individuals, it has also been calculated for groupings as large as countries (Ball, 2005).

Calculating the h-index

The h-index can be calculated manually using free internet databases such as Google Scholar, although subscription-based databases such as Scopus and the Web of Knowledge provide automated calculators. However, one must be aware that each database is likely to produce a different h-index for the same scholar, because of their different coverages summarised below:

- **Google Scholar**
  Google Scholar contains more citations than the other two, but their collections tend to be more accurate. Meho and Yang (2007) found that it had the best coverage of conferences and most journals, but had only limited coverage of pre-1990 publications. It has also been criticised for including 'grey literature' in its citation counts (Jacsó, 2006). However, Meho and Yang's study showed that the majority of the additional citation sources it uses are high-quality refereed sources.

- **Web of Knowledge**
  Meho and Yang (2007) found that Web of Knowledge had a strong coverage of journal publications, but poor coverage of conferences, a particular problem for those disciplines in which conference papers feature significantly, such as computing and engineering.

- **Scopus**
  Meho and Yang (2007) found that Scopus covered conferences well, but had poor coverage of publications prior to 1992.

Sanderson (2008) has suggested that in order to deal with the variation across different databases, one can assume false negatives are more problematic than false positives and use the maximum h-index measured for an academic. Because the content of all databases, particularly Google Scholar, continually changes, any citation research on the content of the databases dates quickly and a scientist's h-index needs to be recalculated regularly.

Pros and cons

The h-index has shown many advantages, but can also be a mixed blessing. Some of its pros and cons are as follows:

Pros

- It can address the disadvantages of other bibliometric measures, for example ones based just on the total number of papers or the total number of citations. The former does not account for the quality of the individual's output, while the latter can be disproportionately affected by participation in a single influential publication. The h-index simultaneously measures the quality and sustainability of scientific output, as well as, to some extent, the diversity of scientific research.

- It is much less affected by methodological papers proposing new techniques, methods or approximations, papers which can be extremely highly cited.

- It serves as an alternative metric to journal impact factors.

- Hirsch has demonstrated that the index predicts high honours well, such as membership of learned societies or even being a Nobel laureate. In his chosen field of physics, a moderately productive researcher will have an h-index equal to the length of their research career, whereas in the life sciences, particularly the biomedical sciences, the h-index tends to have a higher value.

IMPACT FACTORS: THE H-INDEX (CONTINUED)

Cons

There are a number of situations in which the index may provide misleading information about a scientist's output (Wendl, 2007). They include the following:

- It tends to work properly only when comparing scientists working in the same field, because citation patterns differ widely in different fields.
- It disadvantages scientists with short careers, regardless of the importance of their work.
- It doesn't consider the context of citations. Citations are often made simply to flesh out an article's introduction, having no other significance to the work.
- It doesn't resolve other contextual problems such as citations being made in a negative context or citations to fraudulent or retracted work.

Also, it does not account for confounding factors. These include the following:

- 'Gratuitous authorship', which is still common in some research cultures. This is the so-called 'Matthew effect' (the rich get richer and the poor get poorer) and the favourable citation bias associated with review articles.
- It has been found to have slightly less predictive accuracy and precision than the simpler measure of mean citations per paper (Lehmann, Jackson and Lautrup, 2006), although these findings are contradicted by another study (Hirsch, 2007).
- The index produces a natural number and thus lacks discriminatory power.
- While it does not emphasise singular publications in favour of sustained productivity, it may do so too strongly. Two scientists may have the same h-index value, say 30, but one has 20 papers that have been cited more than 1,000 times and the other has none. Clearly, scientific output of the former is more valuable. Several solutions to correct this have been proposed, but none has gained universal support.
- It doesn't account for the number of authors of a paper, which could allow deliberate distortion. Two equally capable researchers might agree to share authorship on all their papers, thus increasing each of their h-indices. Even in the absence of such 'gaming', the h-index tends to favour fields with larger groups, for example experimental over theoretical. A suggestion is to have an individual h-index normalised by the average number of co-authors (Batista, Campiteli and Kinouchi, 2006).
- Most importantly, the index is affected by limitations in citation databases.

Andy Barrow
Faculty Librarian (Science)
Ext.3236
andy.barrow@port.ac.uk

References

Ball, P. (2005). Index aims for fair ranking of scientists. *Nature*, 436, 900.

Batista, P. D., Campiteli, M. G. and Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179–189.

Hirsch, J. E. (2007). Does the h-index have predictive power? *Proceedings of the National Academy of Sciences.* 104 (19193–19198).

Jacsó, P. (2006). Dubious hit counts and cuckoo's eggs. *Online Information Review*, 30(2), 188–193.

Lehmann, S., Jackson, A. D. and Lautrup, B. E. (2006). Measures for measures. *Nature*, 444(7122), 1003–1004.

Meho, L. I. and Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs. Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125.

Sanderson, M. (2008). Revisiting h measured on UK LIS and IR academics. *Journal of the American Society for Information Science and Technology*, 59(7), 1184–1190.

Wendl, M. (2007). H-index: however ranked, citations need context. *Nature*, 449(7161), 403.

WEB 2.0: WHAT'S RESEARCH GOT TO DO WITH IT?

You've probably come across the term Web 2.0 and you might even use some of its tools for your private or social lives. But you may not have considered that such web resources could be useful in your research for resource location, contact discovery, and dialogue and debate with others working in the same field. After all, we're not supposed to be stuck in our ivory towers isolated from the 'real' world and working in a vacuum. If research includes locating the right piece of information at the right time, and engaging with peers to refine and evaluate the work we're doing, then Web 2.0 has a lot to offer.

It's not so much a particular technology as an attitude of mind. Examples are as follows:

- **Photo sharing** sites which allow tagging and rating and commenting on pictures
- **Blogs** which allow dialogues to start through comments on short diary-like entries
- **Bookmarking** sites where users can share their web bookmarks with others and more easily find connected sites of interest
- **Wikis** which enable collaborative working and shared information banks

Tim Berners-Lee, the father of the World Wide Web, commented that Web 2.0 is much more what he initially conceived as the web, 'The original thing I wanted to do was to make it a collaborative medium, a place where we can all meet and read and write… Collaborative things are exciting, and the fact people are doing wikis and blogs shows they're embracing its creative side.' (Carvin, 2005).

What is Web 2.0?

Initially the internet and then the World Wide Web tended to be a great means of publishing; academic, personal or corporate. Web 2.0 signifies more interactivity and a dialogue between users rather than one-way communication. Typically, Web 2.0 applications allow 'tagging', user-generated keywords which allow searches for items with the same tag. They may allow 'comments' and they can sometimes have 'rating' systems (think of Amazon where you can not only rate the books it offers, but also the quality of the comments people make on books). It's sometimes known as the read/write web.

So, how can we embrace its creative side in our research here at the University of Portsmouth? We encourage you to explore the tools which work for you and appeal to you, but here are a few current ones:

Bookmark sharing

You may well be using your internet browser to bookmark websites of interest and relevance to your research. Sites such as Delicious, Diigo, Stumbleupon and many more, allow you to not only share these with other people and even annotate them, but better yet perhaps, see what bookmarks they are collecting, which might well lead to sources you'd not discovered previously. CiteuLike (www.citeulike.org) – 'a free service for managing and discovering scholarly references' and Connotea (www.connotea.org) – 'free online reference management for all researchers, clinicians and scientists' are particularly favoured by the research community as you can see from their straplines and are sometimes called social citation services. The serendipity of finding related things that others are interested in which you'd not previously discovered yourself can be extremely powerful.

Photo sharing

At first glance this might appear to be just a hobbyist sort of thing, but depending on your research, photo sharing can be a way of showing people what work you're doing. Geography field trips would be a good example of this. Alternatively, it might even be something as simple as pictures of a conference or seminar you've attended which can help to build community in your research field. You might even go further and use YouTube to share videos of your work, your research or experiments and so forth.

Blogs

Blogs can promote your work, keep an online log of your research (easily searched by yourself, or if you allow, others), or enable those interested in your research to keep up with your developments. Allowing comments on your blog can provide feedback on your thinking. The blogging community can be very swift to engage in discussion and it can be an effective way of finding out what others are saying about your work, whether they're good or bad comments. Also, if you maintain your own web page (or your department does) you can embed an RSS (Really Simple Syndication) feed of your blog into the page to keep it fresh with your latest news (see box above and article on page 5).

In turn you may be following blogs in your research area or searching the 'blogosphere' for sources and information to feed into the research process. One way of following lots of blogs easily is to use a blog reader, which brings all the blogs (or RSS feeds) you've chosen to follow into one front-end and alerts you to each one that has new content since you last visited. Bloglines and Google Reader are two popular and simple tools which allow this. Searching for blogs, or searching for content on blogs, is also very straightforward using tools such as Google blog search (www.blogsearch.google.co.uk) or Technorati (www.technorati.com). These can give very up-to-date results just hours or even minutes old.

RSS – Really Simple Syndication

This is a way of delivering regularly changing web content such as news and blogs. If you subscribe to an RSS feed, you can get the latest headlines or postings in one place as soon as it is published, without having to visit the individual site from which you've taken the feed. You will often see the symbol above marking a feed that you can subscribe to.

Wikis

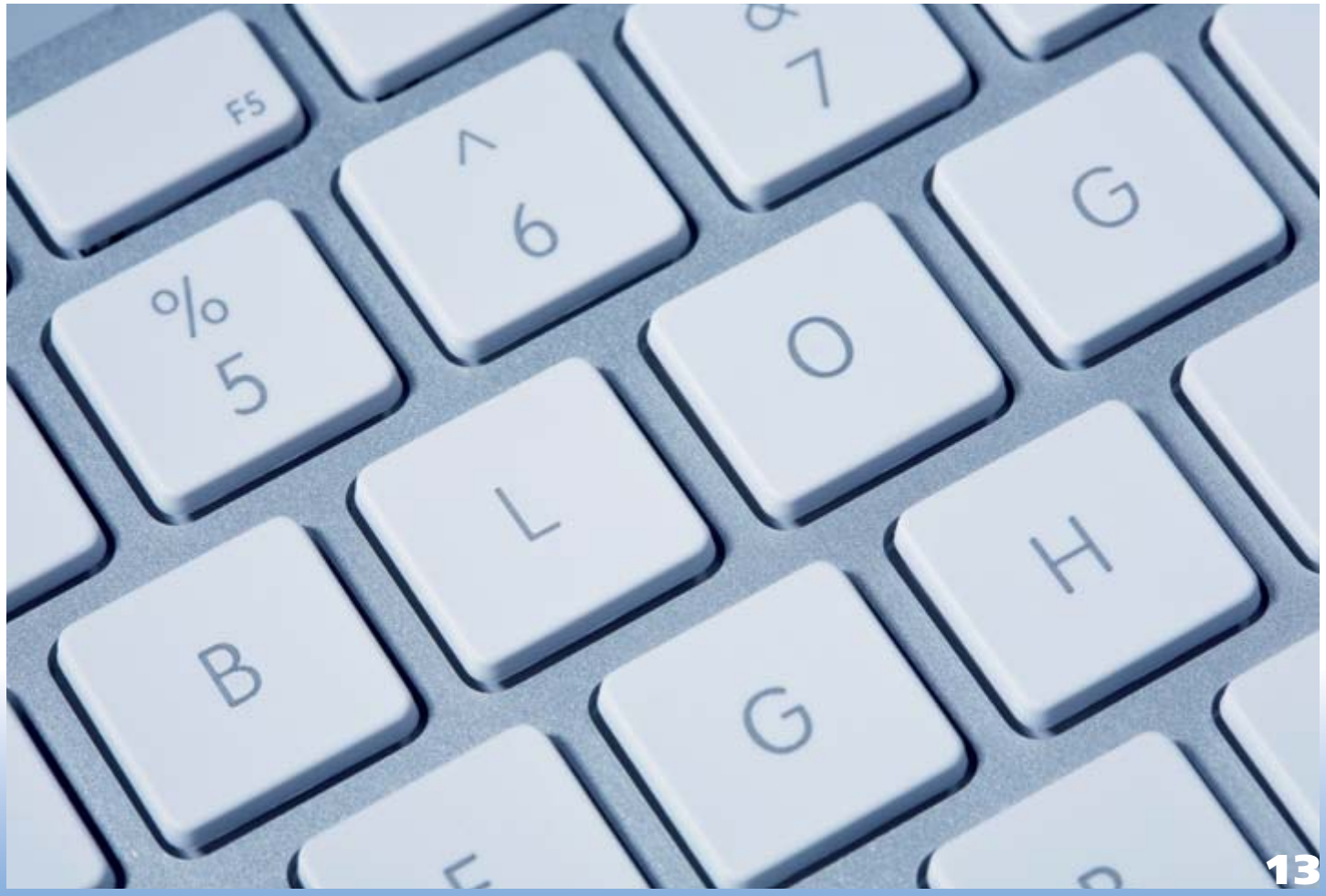
Wikis are free information resources anyone can edit. They are an often misunderstood means of collaborating. A group can use them to develop a set of shared online resources (data, procedures, reference material) with features that a shared hard drive doesn't allow: commenting, revision tracking and simultaneous working. They can be used to share notes made at a conference or for co-authors to work on a paper together. Google Docs allows this kind of collaboration as well. Using either tool is far simpler than emailing a constantly varying Word document between several writers. One of the best examples, Wikipedia, often dismissed as a serious academic tool, can have its uses in providing jumping-off points in the reference lists that many articles provide.

Microblogging

Exemplified by Twitter, it's a very short form of blogging, but it can be a powerful tool connecting you to others in your field or with similar interests and allows very fast communication of the latest developments. The sheer wealth of data on Twitter can also be mined and Ben Parr (2009) offers '5 Terrific Twitter Research Tools'. If you're attending a conference, try finding out whether there are 'hashtags' assigned to the conference, so that everyone who Twitters using that hashtag can see what others at the same conference are saying and thinking. The Library hasn't yet come across really good examples of this being used directly for research – so be the first and let us know what you do and how you do it. In the meantime, you might find it helpful to follow our Twitter feed to keep abreast of Library developments www.twitter.com/uoplibrary.

Virtual worlds

Virtual worlds, such as Second Life, may seem a very unlikely addition to your belt of Web 2.0 tools – unless of course your research



Facebook, Twitter, YouTube: Web 2.0 why go there?

Friday 14 May 2010

The Dark side of the web: how to protect your privacy and security

Friday 30 April 2010

concerns Second Life itself. However, as in the real world, virtual worlds can provide enormous opportunities for locating people and resources, exploring events and geography – but this time without leaving the comfort of your own desk. Search in virtual worlds is still very primitive, so it can be hard work finding what you need, but the rewards can make it worthwhile and of course, word-of-mouth and networking can be great ways of interacting with people and making new contacts. It is even possible to use inhabitants of Second Life to conduct research – although obviously this should never be done without their consent and the usual ethical considerations of any research out in the real world. Walton and McDonald (2009) offer notes on the ethics of using Facebook, for example, which would be applicable here as well.

Training

Training in using Web 2.0 tools can be helpful and there is much around the University that can assist in this – possibly in your own department or faculty. The Library offers workshops for staff which you might like to consider attending (see box above) – or invite us over to run a session just for your research group.

Web 2.0 – why not?

Alison Williams of Southampton Solent University writes (personal communication, November 26, 2008) 'What you can do with these and all the various Web 2.0 tools is very much up to you to explore and create. It's like life, or even Second Life, in that it's very much a blank canvas and a set of tools just lying there, and what it becomes depends upon the participants to make of it what they will.'

It used to be that using technologies like this would require huge inputs on the technical side from computer services and possibly a long lead time while SWOT (strengths, weaknesses, opportunities, threats) analyses were carried out, budgets managed, training arranged and the technology implemented. Now, however, as many of these applications are delivered directly via the web rather than needing software to be downloaded to a local computer, there's every possibility of simply jumping in and testing the water as it were. Indeed, even as long as two years ago Henry (2007) was able to write about how sites such as Facebook make many of these technologies mainstream.

However, it's worth being aware of the risks that such services can disappear or change name and are outside the control of the University. An example of this is Furl which first appeared in 2003, but was shut down in April of this year having been bought out by Diigo. Privacy and security issues are important to bear in mind as well – hence our workshop on the Dark Side of the Web. There's often no easy way of deleting your footprint on the World Wide Web, so it's worth bearing in mind that anything you 'publish' in words, pictures or video form could be around for a long time afterwards. There might also be issues related to user comments, particularly if you choose not to moderate them. There are risks, but there are 'potentially greater risks in failing to engage with a rapidly changing environment.' (Kelly et al, 2009).

With the wealth of available resources and tools, there's every motivation to find those which suit you and your particular field of research and make them work for you.

Timothy Collinson
Faculty Librarian (Technology)
Ext.3224
timothy.collinson@port.ac.uk

References

Carvin, A. (2005). *Tim Berners-Lee: Weaving a Semantic Web.* Retrieved 14 December 2009, from http://www.eioba.com/a1192/tim_berners_lee_weaving_a_semantic_web

Henry, A. (2007). *How Facebook is bringing Web 2.0 mainstream.* Retrieved September 4, 2009, from http://www.mappingtheweb.com/2007/07/19/facebook-web-20-mainstream/

Kelly, B., Bevan, P., Akerman, R., Alcock, J. and Fraser, J. (2009). *Library 2.0: balancing the risks and benefits to maximise the dividends. Program: Electronic Library and Information Systems*, 43(3), pp311-327.
[Also available online in Word and PDF format in the University of Bath's Online Publications Store: http://opus.bath.ac.uk/15260/]

Parr, B. (2009). *5 Terrific Twitter Research Tools.* Retrieved September 4, 2009, from http://mashable.com/2009/05/03/twitter-research-tools/

Walton, N. and McDonald, C. (2009). Facebook and research. Retrieved September 4, 2009, from http://open.salon.com/blog/researchethics/2009/06/14/facebook_and_research

MD4728 1209