CURRENT SCIENCE

Volume 89 Number 3 10 August 2005

EDITORIAL

Institutional E-print Archives: Liberalizing Access to Scientific Research

What do scientists want? 'To be left alone, so that they can think, experiment, discover and invent new things generally (but not always!) for the benefit of humanity' thinks the Common Man – or at least, that is what we, the community of scientists like to believe. The more cynical (and perhaps more realistic) guess is 'money, power, fame and opportunities of self-advancement (though not necessarily in that order), like any other professional group. A more 'scientific' approach would be to find out what scientists actually like, what makes them happy and unhappy. It takes just a little deeper digging into the day-to-day life of active scientists in an academic environment to see the one common source of joy (or sorrow) – the acceptance (or rejection, unfairly of course) of a manuscript in a prestigious international journal.

For the more seasoned of the researchers, the joy is followed by the realization of the long road ahead. Though the change from 'manuscript in preparation' to 'accepted for publication' is a giant leap for the work, it is merely a small step towards its eventual recognition as an important advance in the field. Appearance in print of a good piece of work actually means that the painful task of increasing the visibility of the finding has just begun. There are (or used to be) pleasures too; the arrival of a reprint request letter in the morning mail, the growing satisfaction as the pile of such letters grew larger, the excitement of finding one's paper actually cited by someone, and so on. As the Citation Index gradually gained popularity as a tool for performance evaluation, however, this excitement slowly turned into anxiety if the rate of citation was not high enough (and, of course, it never was).

For the not-so-famous majority (to which most of us belong), whether the article is cited or not crucially depended on the unknown and uncommitted researcher working in the same area coming to know of it (friends would any way cite you, and the opposition would any way refrain from citing you). For this to happen, the journal needed to belong to the elite club of those covered by the (then all-powerful) *Current Contents*. If not, it should at least be covered by the leading abstracting services. Otherwise, there was very little that even a very conscientious author could do if he or she could not get to know about your valuable article. It therefore became all the more necessary to be able to publish in such high profile journals, and almost everyone is painfully aware of the obstacles that one has to surmount for that. One of them, relevant to the present discussion, is the access to latest research findings. The high (and rising) cost of major research journals put most of them beyond the reach of many. How can one ever hope to carry out research good enough to be published in top-class journals *without* reading those journals in the first place? Well, of course you can, if you are Srinivasa Ramanujan. Since none of us are, the problem of lack of access to current scientific literature becomes critical indeed.

Enter the Internet, the World Wide Web and the Google.

Suddenly, some journals became easily and freely accessible by anyone using a web-browser.

Suddenly, the playing fields are levelled – or at least the dice becomes much less loaded.

Suddenly, the powerful, specialized web crawlers and search engines begin bringing a huge collection of articles literally to the fingertips of the researcher. Just enter a few appropriate terms in the search bar and click the 'search' button and in a jiffy, a long list of 'relevant literature to be cited' is ready (never mind an occasional stunner like locating an article on the palatability of rasagullas when searching for voids in protein structures – to quote from a personal experience).

In short, there is now a much better chance that an article published in a not-so-famous journal will be easily picked up as a relevant reference by a web-searching researcher, if (though a very big if at that) the journal is freely accessible on the web. All other things being equal, this would make the 'open access' journals more attractive to the authors compared to the non-open access ones. The number of open access journals, though steadily on the rise, continues to be small, however, amounting to less than a few per cent. Unsurprisingly, many (though fortunately not all) of the non open-access journals are from commercial publishers. Since the apprehension that opening up the access would bring down the circulation figures does sound very legitimate (notwithstanding the experience of a few who saw an increase in the circulation figures after becoming open access), it does not seem very likely that there would be major policy changes in this regard.

Many of the prestigious journals from the commercial publishers (and a few even from learned societies) have a high price tag, and many of the libraries, especially (but not necessarily) from the developing countries simply cannot afford subscribing to them. This further limits the reach of the articles published in them. Unfortunately (or fortunately, depending on how you view it!) this also limits the number of potential citations of these articles.

Thus, the dilemma faced by a researcher is: whether to opt for one of the rather few open access journals and reach a wider readership via the internet, OR whether to choose for the larger number of non-open-access journals and be resigned to the reduced reach.

The way out of this (somewhat artificial) quagmire is the brilliant and breathtakingly simple solution: Open Access through The e-Print Archive, whose grand goal is 'ensuring unrestricted toll-free full-text online access to the entire refereed research corpus – 20,000 journals, 2,000,000 articles per year', as succinctly expressed by the BOAI (Budapest Open Access Initiative – see http://www.soros.org/openaccess/read.shtml). In fact, 'having been swayed by the emotion myself', I would strongly recommend to anyone who is interested (and especially to those who are not), a visit to the website www.eprints.org, to know more about e-prints in general, to glimpse the scholarly, incisive, passionate, dignified and vigorous debate on the many facets of the issue, to see the very persuasive and fascinating statistics, and to feel the excitement of the e-prints movement that is set to revolutionize the way the scientific literature is searched and accessed. The e-print revolution simultaneously solves two of the major problems faced by the developing nations: improving the visibility of their research and improving access to the research articles from the developed ones.

In brief, self-archiving is just putting up the full text of your article, published in a refereed journal, on a web site easily accessible to all. In a way, this is a part of the natural progression that (must have) started before Gutenberg (where scientists painstakingly copied their manuscripts and mailed them to peers – though this makes it a preprint, pun unintentional). Then there were the 'reprint' days where (resources permitting) authors mailed copies of their articles to those who requested them (and for Ph D students, the joy of getting a reprint from a Nobel Laureate was exceeded only by that of getting a reprint request from a Nobel Laureate!). The advent of photocopying machines made matters even simpler. It was no longer necessary to pay through the nose to the publishers for copies of reprints (the scale of operation being small, and the purpose being academic, copyright violation was not an issue of any significance at this stage). However, you still had to mail them, and to do that, you needed to know whom to mail to, to begin with! The 'unknown' readers were still beyond reach.

Tim Berners–Lee's invention of the World Wide Web has changed all this, for ever. The more ambitious and enterprising of the researchers began to put their publications (not just lists, but the documents as well) on their web sites. The Google and other search engines took care of making them visible to those who cared to look for them. For the majority of the scientific community, unfortunately, the task of taking this up was too daunting. Consequently, the vast majority of the published articles continue to be (rather unnecessarily) invisible and inaccessible.

This is where the e-print movement comes in. In the finest traditions of the open source software movement, the proponents of the mission have produced and made freely available an easy-to-use and efficient software package that takes care of all the problems. All you need is a computer connected to the internet, the files of your publications, and the bibliographic information (journal title, author, etc.). Once this information is keyed in (using the friendly interface) and the files of the articles are copied on it, they become not only easily accessible to the rest of the world, but easily and efficiently searchable as well (visit eprints.iisc.ernet.in to get a feel of it). Further, such e-print-servers can very efficiently communicate with each other – and this interoperability not only greatly increases the visibility of the articles, but allows the users of such archives to easily access articles from other servers too.

What about Copyright Issues? Of course they need to be looked into. Fortunately, however, publishers responsible for more than 80% of the published articles permit self-archiving of e-prints in some form or the other (see www.eprints.org), and the figure is rapidly climbing.

The ideal scale for an e-print server (though this is a topic of endless and often acrimonious debate) is a research institution/university. Imagine a moderate sized academic institution, say with about 200–250 researchers. Even of moderately active (i.e. even without raising the bar), an output of about 800 to 1000 publications per year is likely. A full text archive accommodating this would take up at most 2–3 gigabytes – and the entry level personal computers with their 40 Gb disks would be more than adequate to accommodate the publications of a decade.

It should by now become obvious that starting and filling an institutional e-print archive (containing the peer reviewed publications from the institution) is easy, inexpensive, and immensely beneficial to all – a truly win-win-win situation. That automatically ensures that a majority will be totally apathetic towards it, at least in the beginning – such is the human tendency. However, the higher powers (the funding agencies, especially the public/governments ones, though thankfully not yet in India) are beginning to see the advantages of such archives being set up. This is the surest way of truthfully declaring that the results of publicly funded research (at least, in the form of peer reviewed scientific publications) are indeed accessible to the public – should it be interested. An unintentional consequence of setting up an e-print repository is that it immediately brings to limelight the productivity of the institution. Then it becomes a matter of prestige to make the 'score' higher than the competing institutions – and a healthy competition is set in motion. Suggestions about making institutional self-archiving mandatory are being made (by scientists themselves!) with increasing frequency, and one hopes that it will not come to that. One finds photocopying machines, personal computers, mobile phones, colour printers and scanners in every laboratory – and nobody ever made any of this mandatory. Similarly, institutional e-print repository addresses a strongly felt need, and one is bound to see a steady growth of such repositories in the country and in the world in the days to come. What can be a better testimony (and foundation) for the progress of science than the fact that the entire corpus of scientific and technical knowledge is at the fingertips of everyone – all answers just a click away. That is e-print revolution for you.

N. V. Joshi