

Public funding, public knowledge, publication

Based on a paper given at the 26th UKSG Conference, Edinburgh, April 2003

The argument that publicly funded research should be publicly available is a valid one, but it cannot be the most important one in the discussion as to whether research should be freely accessible. The overriding argument is that freely accessible research optimises the scientific process as well as its 'translation' into societal benefits. Free access, or 'open access' as it is widely called, can be brought about by making full use of the technologies available to the world, particularly the internet, but it does need a change in traditional economic models of publishing.



JAN VELTEROP

Publisher

BioMed Central Ltd

Introduction

Having to pay tax is an almost universal phenomenon. This article addresses the issue of access to publicly (tax) funded research. That may not be the majority of research, but it is a sizeable proportion, although figures published (e.g. by the OECD) are difficult to compare country-by-country and make no difference between published research and unpublished research when comparing public funding with privately funded industrial research and development. Much of the latter is unpublished industrial development. It may well be that of the published research, the majority is publicly funded. Some funding comes from private charities that intend their financial support to benefit society, and therefore should perhaps also be regarded as public funds.

Scientific research is one of the things paid for out of tax revenues. Yet tax revenues also pay for the right to read the results of that scientific research, as much of the funding for universities and their libraries (and their subscriptions to scholarly journals) usually comes out of the public coffers. Should any publicly funded research become public knowledge? Indeed, should any information gathered with the help of public funds be public? The question looks fairly straight-

forward, but not all is what it seems. Imagine the information gathered by MI5 or the CIA or Interpol or the Taxman. All done with public money, but nobody would expect that information to be publicly available. In fact, we would expect that information to be highly secret. Indeed, society benefits most from the resources spent and the information gathered if the latter remains secret, at least until those who need to be caught, have been.

But if we narrow the question to scholarly research funded from the public purse, what would the answer be? It would be fair to expect the resulting knowledge to become publicly available. After all, the main reason public money is used for research is that knowledge gathered in that way is used to build the edifice of insight and understanding, for the benefit of society at large.

The argument that scientific results should be freely available because it is tax money that paid for the research is flawed, of course. It is a good argument, but cannot be the defining one. The Americans haven't paid for the scientific results that come out of France; the British haven't paid for research done in the USA, and so forth. Tax is national, but science is global. The far more powerful argument is that the nature of science

requires intensive communication. The benefits of an exchange of knowledge grow with the extent of the exchange. The extent of the communication matters. The internet would not be worth a fraction of its societal value if only a few people had access to it. Its growth increases its value. Communication is the connective tissue between the world's minds. Research results need to be 'broadcast' in order to reach as many of those minds as possible. A scientific mind is, after all, not only the mind of a scientist working in a Western institution with access to all the relevant literature.

Because an exchange of knowledge is an essential ingredient of science, research is almost by definition supposed to be made public. It is expected that it should be published in a scholarly journal. Without such a publication the research is basically deemed not to have taken place. The act of publication constitutes making public.

Public and available

Can we assume the question to be settled? Is there agreement that research results must be publicly available?

Of course they are public. But are they available? If I have a medical condition, do I have access to the results of research into that condition, as published in the medical literature? After all, it's published, that is *made public*, isn't it? Well, not exactly. Publicly available does not always mean freely available. Being published in the scholarly literature is not necessarily the same as being accessible, just as the freedom to buy an expensive sports car doesn't mean everybody is in a position to go and order one.

It would be good to examine what happens in the traditional scholarly publishing model. Researchers do their research, write up the results in an article, and submit it for publication in a scholarly journal. The publisher or the editors of that journal organise some form of peer-review, and if the article is deemed to be appropriate and of the requisite quality standard it is accepted for publication. The authors are notified and then required to transfer their copyrights in the article, or at least the exclusive publication rights, to the publisher. In many cases they are also required to pay the publisher an amount of money, especially if the article contains colour plates. If that's all done, the publisher proceeds to publish the article,

nowadays often online as well as in print. Subsequently, (access to) the article is for sale. And everybody is free to buy it.

Realistically, though, what does that mean? If one wants to know more about a medical condition it is unlikely that the need or desire to know will be satisfied by finding only one article about it. If anything, one will probably look for, and find, more than it is possible to cope with. So it may be a better idea anyway to go to an expert for a considered opinion rather than trying to analyse the plethora of information oneself. The expert will have access to all the information.

Unfortunately, the expert is quite unlikely to have access to all the up-to-date information, either. He or she probably has access to a relatively large proportion of the relevant information, but there is not a single university or research centre that is likely to have access to everything that is relevant to their faculty or research staff. That would just be too expensive. Not even the 'Harvards' of this world could afford it. The problem is eased somewhat by arrangements like interlibrary loan and document delivery, but those remain on the whole cumbersome and fairly slow processes, particularly because it is often prohibited to transfer electronic copies of the articles. And they are not cheap, either.

Common knowledge

The question is, is it a problem? It is, because not only is the value of science greater if it is seen and carried out as a collective enterprise, but scientific knowledge and understanding is a bit like the atmosphere. It belongs to us all. It's our common human heritage. It's the fuel for much of what we call progress. But let's leave aside the access of Everyman to science. The more pertinent issue is that science itself is seriously sub-optimised by the primitive way we still communicate. It's not as though the technology isn't there. The internet makes a massively more efficient mode of communication easily possible. It is the way the traditional publishing models inhibit the internet's potential that is the stumbling block.

Analogous to open source software¹, science can be thought of as 'open source knowledge'. That is how science would work at its most optimal. All the knowledge being freely accessible and then extended, built upon, improved, verified or

falsified, discussed, applied, accepted or rejected, etc., mostly by scientists, occasionally by others. The traditional publishing model that by its very nature limits access to the knowledge that could be available is not suitable for this 'open source' approach. To some, the idea that scientific knowledge should be available to anyone outside scientific circles is anathema. They argue that it is sufficient if the information is accessible by other scientists who know what they are doing with it. This is arrogant and patronising, and assumes that scientific information does reach all those who matter. If that would be the case, it is difficult to explain the strong growth in accesses and downloads when articles are made more widely available, even if just in the big deals. But it isn't difficult to think of scientific journals for which this arrogance is an integral part of the brand. In their defence, the ivory tower attitude to publishing is exactly how things were for a very long time and, on the whole, scientific research has been carried out successfully, though sub-optimally, in that way for decades, indeed centuries. The thing to realise is that there was no practical alternative. Printed journals were the means of disseminating knowledge, and the physical distribution of printed matter combined with the cumbersome ways of comprehensively finding the right information naturally limited access to it to the initiated scientific classes. And even they rarely have anything approaching comprehensive access to the science literature. So the 'open source knowledge' idea doesn't even fully come into its own in scientific circles.

Interests and money

Before the internet era the traditional subscription model for publishing science was perhaps not ideal, but the best possible. However, it came with limited circulation and expensive distribution. But now there is an alternative. Now, electronic publication makes virtually limitless circulation possible, at least to anyone connected to the internet, and the cost of distribution is very small indeed.

Going back to the scholarly publishing model, publishing obviously does cost money. Cost elements include the organisation of peer review, editing, coding (SGML, XML) for the web, preparation for print, print and postage, web hosting, web access-control, marketing and promotion,

sales, customer service, and the cost of inefficiencies such as different sizes and standards for different journals. Even 'brand-creation' has been mentioned, although that is in my view a function of marketing and the real 'brand', the 'prestige' of a journal, is in many ways the creation of its editors and the authors publishing in it. The traditional way to recoup the costs is for the publisher to require the authors to transfer their copyrights and then sell their articles via subscriptions to his journals. Conventional publishers are little more than 'copyright-mongers'. Copyrights, or those components of copyrights that could be described as 'exclusive distribution rights', are key terms in the equation, as they create artificial scarcity of the articles and without these rights the publisher has little to sell in order to recoup his costs. The ideal copyright line for a publisher is: © *The Publisher. No copying or further dissemination of this article is allowed.*

But what about the author? An academic author writing up research results is not the same as a fiction writer. Such an author doesn't make money by writing, but by doing research, and writing up the results is like keeping the minutes of a meeting; it is part and parcel of the work done. It has been argued that scientific research articles should be freely available because they are 'given away' by researchers to society (many sources, e.g. Stevan Harnad²). This sounds very noble, but it's not quite to the point. After all, authors need to publish ('publish or perish') and in reality the articles are not given away to society, but to publishers. Actually, they are not 'given away' at all, but 'sold' in exchange for the currency that science craves: the 'label' of acceptance in a journal and the associated credibility in terms of the journal's prestige. Joshua Lederberg once put it like this³: "*Their gain from publication is recognition by their colleagues and the dissemination of knowledge in the spirit of science.*" Harnad⁴ has called it '*The Faustian Bargain*'. The ideal copyright line for the author, therefore, is different from the publisher's. Authors need to be cited, and in order to optimise the chance to be cited, they need to be visible. That is imposed upon them by the scientific career and reward structure. It's not just 'publish or perish', but also 'be read or be dead'. An author's motto pretty much must be "I am cited, therefore I am". So the copyright line an author wants is this: © *The Author. Please copy and distribute this article as often*

and as widely as possible. Indeed, the latter is not only the ideal copyright line for the author, but also for the public funding body, whose charge is to fund relevant research, for the benefit of science and society at large.

From here to there: the transition

Clearly, there is a conflict. Maximum distribution, the ideal situation for author and funding body alike, is not compatible with a subscription model. 'Big Deal' access licensing models potentially increase circulation, but they are fundamentally just attempts to shoehorn electronic publishing into the subscription model and the 'bundled subscription' at that. New models are needed to provide the funds to recoup the publisher's legitimate costs. It may be worth looking at the way advertising works. The very idea of advertising is to reach an optimal number of the right people. Isn't the purpose of communicating scientific research similar? Imagine advertising being made freely available to publishers who then sell access to it. It is not so easy to see how that would work. But that is exactly what's currently happening with scientific research material. Is the crucial difference that people aren't much interested in advertising but they do want to read science? I don't think so. It could be argued that many advertisements contain interesting, relevant information if you are looking for what they offer. And conversely, many scientific articles are not relevant to your research and therefore not of interest.

Just as with advertising, the goal of reaching an optimal number of the right people is best met by a publishing model in which the publisher is paid for the service of publishing and initial dissemination of research papers at the input-side of the process. The internet makes such a true 'open access' model possible. Open access not only makes scholarly communication more efficient, it also enables the idea of science as 'open source knowledge' to flourish. And since one of the principal aims of publicly funded research is to benefit society, it is important to realise that open access also optimises the translation and transition of scientific discoveries into societal and economic benefits.

A reminder of what 'open access' actually means*:

1. it applies to articles, not necessarily journals or publishers;

2. the author declares that the article can be used by anyone for any legitimate purpose;
3. the article is archived in a suitable format in at least one internationally recognised online open access repository (such as PubMed Central).

The big question is not so much 'should there be open access to scientific knowledge', as the answer to that is an increasingly clear 'yes', but rather, 'how do we get to open access from the current situation of very limited access to this knowledge'.

If we define 'open access' as we have done above, it points us to a few ways the transition could be facilitated. The fact that open access applies to articles rather than journals or publishers permits a mixed model. That is important for a transition. The amount of revenue an average STM publisher currently makes (so-called 'top-line revenue') from an average article is in the order of \$5000 (according to Andrew Odlyzko⁵ actual amounts varied, when he did his study, between \$1000 and \$8000 per article, dependent on the journal, the discipline and the publisher; my recollection from the time that I worked in the traditional journal publishing world is of a narrower range – between \$3000 and \$7000 per article – and a recent 'ball-park' figure, which I obtained by taking the published or estimated journal revenues from a few of the larger commercial publishers divided by the number of articles published by the same publishers, point to a per-article amount of well over \$4500; this figure is an average and varies substantially for individual journals in the publishers' programmes). If such an amount of revenue per article could be realised from input-charges, at least some publishers would have no problem reversing their business models and converting to open access policies. In the meantime, they may be tempted to offer authors the choice and raise sufficient amounts from individual articles for which the author, or his funding body, insists on its free online availability. It may not be realistically possible for authors to come up with the same amount the publisher makes now, especially not if they are toward the high end of

*A more 'legal' definition and an author's copyright and license agreement can be found on the BioMed Central web site <http://www.biomedcentral.com/info/about/license>

the spectrum, but those amounts cover costs that are eventually redundant in an open access environment, such as the cost of print and access control. A transitional mixed model phase (David Prosser⁶), in which the author is given the choice to pay for open access – or not to pay and accept limited dissemination – may give the publisher some comfort and time to adjust. The stance that an increasing number of funding agencies is taking, namely that the cost of publishing is to be seen as an integral part of the cost of the research itself, is most helpful.

The definition of open access also mentions archiving. The point of archiving articles in open access repositories is that it gives the author reassurance that the article will be properly preserved. It also reassures the publisher that he doesn't have to commit to long-term investments in maintaining his own archives. (Incidentally, the concept of publishers maintaining archives has only been around for less than a decade, since online publication took off seriously; before that, archiving was almost exclusively the libraries' role).

Where does open access stand now?

At BioMed Central we fully adhere to the principles of open access for any research article we publish. Indeed, we have helped to define much of what open access now is. It is not (yet) an easy task. The environment in which we work is still heavily stacked against such radically new ways of publishing scientific results, however beneficial they may be. Not only authors, but the entire scientific community, even society at large, would benefit from open access. Yet, the onus of effecting the change from the conventional publishing model to an open access model is pretty much on the shoulders of individual authors at the present time. The sad fact is that the existing pressures on authors to conform sometimes prevent them from choosing open access. Regarding publication charges as part of the cost of doing research (as an increasing number of institutions and funding bodies do) will help. What will also help is a commitment on the part of funding bodies – and those who evaluate researchers for appointments and promotions – to base their judgement on the merit of articles published rather than just on the journal in which they appeared. The growing recognition that an author's decision to publish

articles with open access must be seen as a service to science is also most welcome.

Apart from a few small independent journals⁷, BioMed Central still is the only publisher to apply the input-paid model at any scale at all. The journals that the Public Library of Science⁸ intends to start publishing later this year will follow the same economic model as BioMed Central. The cost-structure of an input-paid model is very different from the traditional model, for a variety of reasons. First of all, the charges levied by BioMed Central (\$500 per article) and Public Library of Science (\$1500) cover only electronic publishing. Print versions are available at an extra charge. The charges currently apply to published articles only, and so 'subsidise' the review process also for the rejected manuscripts. This may or may not be sustainable and it is very possible that a proportion of these article-processing charges will in future be non-refundable for rejected manuscripts. The charges are able to be kept this low in part because both BioMed Central and PloS can set up the publishing operation from first principles, not encumbered as they are by historical and organic developments typical in an older print publisher, usually resulting in inefficiencies that are difficult to remove.

BioMed Central has devised an option of 'institutional membership' in an effort to lower the hurdles and take the burden of payment off the shoulders of individual scientists. This may turn out to be a temporary arrangement, especially if publication of results is being widely recognised as an essential part of research done, and the cost of publication consequently regarded as part of the cost of doing research and paid for by those who fund the work.

References

1. The Open Source Initiative:
<http://www.opensource.org/index.php>
2. Harnad, S., 'The Self-Archiving Initiative.' *Nature Web Debates* (26 April 2001):
<http://www.nature.com/nature/debates/e-access/Articles/harnad.html>
3. Lederberg, J., Options for the Future, *D-Lib Magazine*, May 1996:
<http://www.dlib.org/dlib/may96/05lederberg.html>
4. Harnad, S., Electronic Scholarly Publication: Quo Vadis? *Serials Review*, Vol.21,no.1, pp70-72, 1995;

- Reprinted in *Managing Information*, Vol.2, no.3, 1995:
<http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad95.quo.vadis.html>
5. Odlyzko, A., The Economics of Electronic Journals, *First Monday*, 1997:
http://www.firstmonday.dk/issues/issue2_8/odlyzko/#III.
 6. Prosser, D., SPARC Europe on Open Access, *SPARC E-News*, February-March 2003:
<http://www.arl.org/sparc/core/index.asp?page=g29#6>
 7. Such as:
Conservation Ecology:
<http://www.consecol.org> ,
New Journal of Physics:
<http://www.njp.org>
 8. *Public Library of Science*:
<http://www.publiclibraryofscience.org>

Jan Velterop
Publisher
BioMed Central Ltd, 34-42 Cleveland Street, London
W1T 4LB, United Kingdom
Tel: +44 (0)20 7323 0323
E-mail: jan@biomedcentral.com
Website: <http://www.biomedcentral.com>
