

An introduction to self-archiving and Open Access literature for independent scholars

Ross W. Gayler (r.gayler@mbox.com.au)

2nd April, 2005

As scholars, we are producers and consumers of research literature. As independent scholars our access to that literature may be somewhat restricted relative to our institutionally dependent colleagues. Fortunately, internet access to the research literature is now more than adequate in many fields, and continually improving. This brief note serves as a guide to some of the internet services and developments of interest to scholars. There is no novel content here, merely a convenient collection of information that is not as widely known as it should be. It assumes that you are already familiar with internet browsing. Please accept my apologies for the bias towards science and technology in the examples.

The spread of scholarly knowledge has always depended on the available technologies for communication. Prior to the advent of printed journals, knowledge transfer depended on physical transfer: face to face meetings, correspondence, and transport of the objects of study (such as botanical samples or sculptures). These modes of communication have some disadvantages. The number of scholars who can simultaneously access the same knowledge is limited by the number of people who can be crammed in a room and there is a significant cost attached to transporting people or objects. Furthermore, face to face meetings and correspondence are point to point communications, which raises the problem of how an outsider learns of the existence of a conversation in order to participate.

The printed scholarly journal was a major advance in that it allowed one author to address a much larger audience (scattered over space and time) than could be reached by lectures and correspondence. It also addressed the point to point problem by providing fixed locations (specific journals) where outsiders could search for material of interest. As the volume of research literature expanded, further mechanisms such as citations and indices were added to serve the same function of providing access points into the mass of literature.

The technology of print publication served the needs of the research community. However, it also brought with it characteristics that are a function of the nature of printing rather than the nature of research. Printing and publishing of physical documents is expensive. These costs are met by journal subscriptions. However, when the increase over time of journal space is combined with the profit growth imperative of for-profit publishers, the net result is a growing economic barrier to research dissemination. The heart of this problem is that the researcher wants to give away results as widely as possible whereas the traditional publisher's business model is based on restricting access to only those who can pay.

This problem has been observed and discussed, primarily from the viewpoint of research institutions, where the issue arises from the finite size of library budgets. Occasionally, it is discussed from an equity viewpoint, where the concern is for researchers in developing

countries. The independent scholar is probably most similar to the latter group in being largely self-reliant and lacking large-scale funding.

The proposed solution to this problem of access is self-archiving: the depositing of an electronic document in a publicly accessible website. This can be done at minimal cost to the author and makes the document available to a global audience at minimal cost to the reader. More generally, the Open Access movement is concerned with removing barriers to readership, and self-archiving is one means to this end. Probably the best introduction to Open Access is Peter Suber's Open Access Overview (<http://www.earlham.edu/~peters/fos/overview.htm>). The arguments for self-archiving have been made eloquently (and at length) by Stevan Harnad, see: <http://www.nature.com/nature/debates/e-access/Articles/harnad.html> and <http://www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm>

The move to self-archiving has been around for some while. The earliest such archive, arXiv, was established in 1991 to serve the needs of the high-energy physics community. In part, it was established to address the tardiness of journal publication. In this field, active research communication is far ahead of the journals, which are seen as publications of record for archival purposes only. A brief history of arXiv is available at: <http://arxiv.org/blurb/pg96unesco.html>

The push for self-archiving has gathered momentum and there are now many organisations providing support individually or through joint efforts. A good central point of reference for an individual contemplating self-archiving is the EPrints self-archiving Frequently Asked Questions list (<http://www.eprints.org/self-faq/>). Research libraries are also supporting self-archiving because of the implications for library budgets (<http://www.arl.org/sparc/about/faq.html>) and a UK Parliamentary Committee has come out strongly in favour of self-archiving and open access publication (<http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm>)

A good example of an Australian self-archiving site can be found at the University of Queensland ePrints gateway: <http://www.library.uq.edu.au/database/eprints.html>. This also provides links to other Australian university archives and useful background information. Independent scholars are likely to be interested in ADT, a database of Ph.D. and Research Masters theses produced at Australian universities: <http://adt.caul.edu.au/>

In addition to institution-based archives there are topic-based archives, such as arXiv for physics (<http://au.arxiv.org/>), cogprints for cognitive science (<http://cogprints.ecs.soton.ac.uk/>), and CiteSeer for computer science (<http://citeseer.ist.psu.edu/>). Since these archives are not linked to specific institutions they provide mechanisms to allow anyone to submit documents.

A more extensive list of free access online archives can be found at http://dmoz.org/Science/Publications/Archives/Free_Access_Online_Archives/ and a current project to establish a comprehensive and authoritative directory of Open Access repositories is described at <http://www.opendoar.org/>.

One consequence of having so many archive sites available is that it can be hard to find the right archive to search in. This has led to the development of cross-archive search services such as Google Scholar (<http://scholar.google.com/>) that are specifically targeted

at scholarly literature, including peer-reviewed papers, theses, books, preprints, abstracts and technical reports. Another example is OAIster (<http://www.oaister.org/o/oaister/>), a search engine that gathers information from many archives with the goal of providing a collection of freely available, previously difficult-to-access, academically-oriented digital resources. This is particularly interesting as it is not limited to documents but also indexes images, movies, and audio recordings. A good use of open access image archiving can be found at the University of Virginia

(<http://www.lib.virginia.edu/rmds/collections/gordon/index.html>) where digital images of rare sixteenth-century printed French books from the Douglas Gordon Collection are being made available to the public:

http://iris.lib.virginia.edu/speccol/gordon/gordonimages/Gordon1547_M35/

In addition to archives, there are also Open Access journals. These differ from archives in having their own refereeing processes. The Directory of Open Access Journals provides a database of free, full text, quality controlled scientific and scholarly journals. The DOAJ aims to cover all subjects and languages. At 30th March, 2005 there are 1,503 journals in the directory at <http://www.doaj.org/>

Traditional journals have varying policies on author copyright and self-archiving. Some journals completely forbid self-archiving, others permit it, and of course there are intermediate positions. A database of publisher copyright policies on self-archiving can be found at <http://www.sherpa.ac.uk/romeo.php>. Advice on dealing with publishers and legally sidestepping copyright restrictions can be found in

<http://www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm#Harnad/Oppenheim>. In particular, it is important to recognise that journal policies on self-archiving are only a contractual matter between the publisher and the author and the copyright pertains only to the specific work published. Therefore, the author is free to self-archive related but unpublished versions of the work (for example, the penultimate draft accompanied by the corrigenda required to convert it to the final version).

Copyright allows the author to reserve all rights in a work. This is appropriate to for-profit publication, where the author wishes to have complete control of the exploitation of the work. However, in the arts and sciences an author may be better served by encouraging others to exploit their work, for example by incorporating it into derivative works. The Creative Commons (<http://creativecommons.org/>) attempts to promote this by developing flexible copyright licences (<http://creativecommons.org/licenses/>) that explicitly permit some derivative uses (for example, sampling of music) while reserving some rights to the author (for example, attribution of authorship).

The move to self-archiving and open access provides an embarrassment of riches although the current coverage is somewhat variable by discipline. The best thing to do is search – you will be amazed at what you can find. If what you seek is not currently available you can wait (because rapid improvement is the norm) or better yet, seek out like-minded people to help you fill that gap.

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.0/au/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.