# Open Access to Scientific and Technical Information:

# The state of the art.

**A report compiled for INIST by Jack Franklin (ASFRA BV)**
**Edam, The Netherlands (October 2002).**

This report has been prepared to highlight some of the issues which are important to the development and adoption of Open Access technologies in the Scientific, Technical and Medical information world. The contents are the responsibility of the author and do not in any way represent a formal position of INIST or the other parties sponsoring the seminar.

## Table of Contents

## 1. Introduction and Scope and Aims of this Report

The advocates for Open Access (OA) aim to establish "common standards whereby articles stored on compliant servers can form a global library allowing searching, data retrieval, cross-linking, and stable long-term archiving "(www.openarchives.org). As such, this points to a technology based initiative, which will open a disseminated, yet interlinked, library of information.
At the same time, many scientists clearly hope that OA, by definition, will offer them the opportunity to change the present ways in which information is stored and accessed. This has led to the initiative having particular emphasis in the Scientific Medical and Technical (STM) sector where discontent about the manner in which information (journal articles) is handled has been rising. Some scientists feel that OA could offer the scientific community a chance to by-pass the conventional publishing process and therefore the publisher. Articles could be stored in self, or institutional, or even "friendly, non-fee-charging, archives", and so eliminate the need for the present scientific journal.

Thus OA has a strong political content; in fact, it is increasingly difficult to discuss the subject without this heavily influencing the tone and outcome of the argument.
Strangely perhaps, OA is also seen by many commercial publishers as a way forward. Market leaders such as Elsevier Science, John Wiley & Sons and Springer all agree that OA targets are useful and achievable. Some of these already offer the user much of what the OA community is looking for, such as linking, multi-file searching, full text retrieval etc. All too often the only difference is how these advances should be paid for.

Given the range of themes this seminar could follow, it is important to state that this initial background report is not meant to be an exhaustive description and evaluation of OA *per se*. Rather, it attempts to sketch some major debating points for the meeting to discuss and elaborate upon. Given that this is focussed on the STM field, the report also looks most closely at the more political side of the discussion – whether or not OA can, should, or will replace the journal in a responsible manner. If this is to be the case then the service has to at least offer acquisition, verification and quality control, production, dissemination (including location and access) and archiving. It must offer a challenge to the present system of scientific publishing and, to be truly successful, allow more to be done with the articles than has hitherto been the case (for instance, to offer text mining opportunities for the better location and retrieval of information).

## 2. The aim of the Open Archives Initiative (OAI)

Many activities surrounding OA technologies are organised under the Open Archives Initiative (www.openarchives.org). It is not worth repeating here the detailed and excellent information available at this address and the FAQ section in particular explains many of the issues that have been raised in the discussions leading to the writing of this report.

Nevertheless, it is perhaps useful to state that the OAI was established in July 1999 in the USA. The plan was, and is, to deal with how document archives, pre-print servers, and digital libraries can be connected and accessed through a common and easy-to-use interface; and to develop related services such as distributed retrieval options, expert information systems and other services.
The initiative clearly has many of its roots in the pre-print, or Eprint, movement and so was initially focussed on the STM sector; although it has gradually spread to other information arenas. The work has been focussed on the public sector but the norms can be, and will be, used by both the free and commercial services. (A good overview of this can be found at http://www.ariadne.ac.uk/issue31/open-archives-forum/).

Supporters of the OAI feel the initiative is necessary due to the increasing amount of information available over the web. In addition to individual articles, whole libraries have placed their catalogues, full text and other digitised resources online and made them freely available. However, there is obviously a (large) amount of heterogeneity in these resources which inhibits the use of all of them under one system.
The OAI has swiftly evolved into an international movement with many active individuals, institutes and organisations. The OAI has two different types of user groups:
1) Data providers who are responsible for repositories and who have adopted the Open Archive Initiative Metadata Harvesting Protocol as a means of exposing metadata about the content of their repositories.
2)Service providers who harvest metadata from OAI compliant repositories and use this as a basis for building value added services.

A key part of the OAI is the Open Archives Metadata Harvesting Protocol. This is only one of a number of such protocols which originated in the library community with Z 39.50, but it is "simpler" and therefore can be used by a larger audience if required. The OAI has recommended that the Dublin Core metadata set be taken as the minimum descriptive level; this has been encoded in a special XML syntax developed by the Initiative. By implementing the OAI protocol, archive, data providers make metadata for their digital contents accessible. External requests to the archive using the OAI protocol are answered in a standardised way.
Service providers harvest metadata made available by the data providers and then add value/services such document delivery services, or specialised search engines.
The services can be charged for, or be offered free. The OAI policy is that any service provider observes the legal policies of the referring archives.

## 3. Eprints, and the OAI in relation to the STM publishing world

A visit to the OAI website will show that there are many initiatives and some technical experiments/solutions underway which are allowing OA to begin. A number of sites are OA compliant. In parallel to the technical advances however, one will find it hard to miss the emotive side of the movement. Many OA supporters clearly see OA as a solution to breaking what they see as "a publisher's hegemony". This is perhaps best illustrated by Stevan Harnad in his

article "For whom the gate tolls". http://www.ecs.soton.ac.uk/ – harnad/Tp/resolution.

OA might, indeed, provide an alternative to the established journal environment. However, to break this present dominance, OA has to offer a solution for authors wishing to put their STM material online. Harnad has therefore put forward some concrete arguments but also solutions, the most significant to-date being the E-prints software (http://www.eprints.org/), written by Harnad's collaborators Robert Tansley and Chris Gutteridge. This offers a web interface for managing, submitting, discovering, and downloading documents. By definition, Eprints repositories are compliant with the Open Archives Initiative (OAI) so that once a collection is registered as an OAI data provider, OAI-aware information services will be able to locate and access its content.

The software appears to be effective (e.g. see the review at http://www.ariadne.ac.uk /issue32/eprint-archives/intro.html). It is also the focus of a long and interesting debate on The American-Scientist-E-PRINT-Forum, http://www.ecs.soton.ac.uk/%7Eharnad/Hypermail/Amsci/date.html#2212.
The Eprints software is available free for academics and non-commercial users from the Eprints server. Early reports indicate that the software is easy to install and use. A further initiative is that Ingenta is also making a commercial version available (http://www.ingenta.com/). Ingenta will also offer customers the possibility of establishing their archive under the company's umbrella as they feel, probably rightly, that many institutions will be unable or unwilling to supply the resources to maintain a full archive with all the security and stability parameters that will be needed.
There are some in the OA community who question the wisdom of this. A commercial software package might, they reason, soon lead to a two-tier system with the commercial software offering more than the free academic version. This need not happen, especially if both partners are determined to keep the two versions of the software completely interactive (e.g. the SRS package has survived privatisation well **[1]** ).

Establishing an archive is one thing but it remains only a starting point. If such initiatives are to replace the publishing process then more added value services are required – from checking that the article is bona fide in the first place, to eventual refereeing and adding of the corresponding metadata for archiving etc. Furthermore, a major function of "publishing" is to ensure/indicate that the materials included are useful. Archiving, i.e. ensuring that materials are permanently available, a lot of "author-preferred" materials might do more harm for the OA cause than good.

Perhaps for this reason, the immediate emphasis is to get authors to archive their scientific papers before they are processed and taken up by the traditional journals. This use of the modern-day pre-print started with the success of the physics pre-print service established by Ginsparg http://arXiv.org/ and has been continued with such services as CogPrints (a site for psychology and biology papers at www.cogprints.soton.ac.uk).

Although the electronic version of pre-prints was new, this idea has been here for many decades, perhaps centuries, as scholars have often sought to distribute

their ideas to a peer group prior to telling the rest of the world. Naturally however, the internet allowed scientists to do far more with their electronic copies of proposed articles than just send them to a few selected individuals. The original pre-print service thus stimulated physicists to submit pre-prints of their papers to a central server which were then accessed and reviewed and were thus available earlier and in a more interactive form than their destined paper journal formats. However, a crucial part of the system was, and is, that the papers would later be refereed and so would acquire the added value of peer review.

The key to any scientific information service is quality. A common and powerful argument from the publishers is that pre-prints, or other forms of self archiving, lack this and so are not adding that much to the information chain. Several supporters of present day OA are banking on the hope that journals will eventually allow the post-print, the refereed version, to be mounted for free alongside the pre-print; but much will have to change before this is the case (although the publishers are in the main willing to allow for links from pre-prints to the post-print as long as it remains on the publishers server – see later).
There is however a strong attraction to the idea of using pre-prints as a new scientific publication. One or two publishers have therefore started to offer pre-print services where authors submit an article to a central server which is mounted in the public domain. The articles may then be read, criticised or commented upon. The author is usually free to decide what to do with the article after this period of "public scrutiny". Some may withdraw the article, others will submit the paper to a traditional journal. In some cases, these services are actually part of the submission procedure to a journal but others allow the author to select another journal after the pre-print period. Examples include Net Prints, in medicine (www.clinmed.netprints.org), run by the British Medical Journal group and Stanford, and the Chemistry Prints Service (www.preprint.chemweb.com) run by ChemWeb (Elsevier).

One should still not think however that the problem is solved. Many journals refuse to accept articles that have been previously submitted to a pre-print server. This refusal has nothing to do with copyright but is just the journal's policy. Supporters of these preprint services claim that the number refusing is small, and, while that is probably not really the case at the moment, a significant number of publishers and individual titles are changing their attitude towards allowing this form of (pre) distribution **[2]** (it is also interesting to note that those refusing are not necessarily the commercial publishers but include societies such as The American Chemical Society that refuses to handle papers that have previously been viewed on the ChemWeb server).

The debate as to why this embargo is in place usually comes down to "copyright and reviewing". Copyright law, while looking basically simple, is a complicated subject (the European Commission is presently working on a Directive which will lead to harmonisation in this area across the Community (http://www.eurorights.org/eudmca/CopyrightDirective.html) and it is to be hoped that this activity will lead to a clearer division of rights than has hitherto been the case). In most cases however a pre-print is indeed free of third party copyright, although some "funders" are enforcing their ownership of the IPR connected with the results of their funding and this can, in fact, also mean they reserve the right to hold the copyright of any resulting articles. Therefore

articles can be self-archived. However, when the pre-print is later submitted to a journal, it is entirely up to that journal whether or not they agree to handle the material.
The reviewing issue is more simple and comes down to what quality one expects from the refereeing process. Net Prints have produced an overview of how they see this developing which clearly explains the situation:
*"Feedback from readers does not constitute peer review as it is currently understood. For authors who wish to submit their work to a peer-reviewed journal, we provide lists of journals that do and do not accept submissions that have previously appeared as electronic pre-prints. In time, we hope to allow authors to activate submissions from this website to the journal of their choice, thus streamlining the peer review process.*
*Once articles have been published by a peer reviewed journal, we would like to direct readers to the full text of this final version, wherever it is. This could be via a hypertext link to the peer-reviewed journal or to the author's personal or institutional website. Authors who want to post the peer reviewed article on clinmed.netprints.org will ensure that they are not signing away this right as part of the licensing/copyright conditions agreed with the peer reviewed journal. We hope to publish journal policies relating to this issue in due course."*

## 4. Acquisition and Refereeing

Few STM articles published today are rarely read thoroughly. The vast majority probably receive no more than a cursory glance (title, abstract, perhaps a run through of the Materials and Methods section and some references). Yet the work was done, it was sound, and it MIGHT be useful to someone. Scientists must still "publish or perish" and so they choose the journal most suited to their subject, and most suited to their work. A standard, routine piece of scientific reporting will find it difficult getting into Science or Nature, or even a top quality discipline specific title such as Brain Research, Yet most articles find a home somewhere. Scientists therefore select the best home they can for their work, and other scientists "pre-select" the importance of an article by the journal it is published in. Overall, a journal's refereeing process is very important – some would say it is the journal's sole claim to quality and later fame.

Many in the OA community are now looking at how to add value/validate/review their materials. Some journals are making their pre-prints OA compliant and then adding a link to the final refereed version. However, as mentioned earlier, some authors do not like their articles being mounted for public consumption until they have been refereed. One example is Behavioral and Brain Sciences - http://www.bbsonline.org/ - which uses Eprints software for its archive and refereeing, but has created separate sectors for those pre-prints whose authors do and do not wish to make it public that their papers are under refereeing by BBS.

There is little doubt that the publishers will support a system whereby the pre-print links to the finished version on the publisher's site. This route does not then remove the problem that many users will be unable, under the present conditions, to access articles they are interested in because they have neither a license nor a subscription (see Section 7). Single article sales would ease this problem, as many publishers, e.g. Blackwell Science, allow. For the others,

there is the small compensation that the user knows that the pre-print version is not the final version although frustration might still be the order of the day.

Another hope is that self-refereeing, reviewing by the community maintaining the server, or the comments from the market, will control many abuses. While the "jury remains out" on the use of pre-prints as part of the publishing procedure, there is no doubt that a considerable body of opinion is in favour of pre-prints being used for "open refereeing". Computer based techniques monitoring the use of online publications are available is some fields and shows that online articles have a higher chance of being cited than those in the paper journals. Such techniques can be used to monitor the evolution of such articles and could give a better indication of usage of papers in the future (for a deeper discussion of this and related issues go to the Nature Debates on E-access: http://www.nature.com/nature/debates/e-access/Articles/).
Some supporters of OA opine that OA will help even if the post-print is never freely archived because the pre-print is in the public domain anyway. If this is a true belief, it is dangerous. Inaccurate information is non-information. A major drawback to the OA initiative will surely be that the reader might not know if he has the last version of the article – computers work very well looking backwards but are less accurate with forward links!

Just as Eprints have spawned centralised pre-print services, it is also possible that players in the OA environment will establish central reviewing services, in different disciplines, which are as trusted as the present ones. It seems more likely though that we will see a number of independent reviewing/validation services. Already some universities are wary of allowing any of their faculties to mount "unreviewed" articles that might put the university in a poor light. Industrial companies have strict rules as to what staff might mount on internal and external servers and so a degree of quality control is already present. In any case, refereeing is so important that OA might well succeed or fail on this issue alone.

To-date, probably due to the short time OA has been with us, few attempts at establishing quality control guidelines for OA sites seem to have been made. The problem is certainly recognised however and so forms of access control and validation as well as content reviewing will surely soon begin. One candidate group for taking action are surely the "Learned Societies". In many ways these could repeat history, they were often instrumental in running the first journals and could now examine whether their members could not support pre-print or similar services in a specialist area. The skills needed to design these services might need to come from parallel initiatives and the emerging expertise in the Digital Libraries field could be an excellent partner.
One further trend might also open new avenues for STM information. Just as the nucleic acid sequence databases "check" data as it is processed, other areas need to deposit data in electronic systems for external checking. OA offers an opportunity for scientists to make their data available for general review and even manipulation. Copyright and "piracy" rules have to be developed and adhered to, but we could soon see the journal article relying far more on data deposited elsewhere.

## 5. Searching, Locating and Linking

There is little doubt that I will be criticised by someone for failing to include some information that I may have missed while researching this report. Yet it is equally certain that without the use of the World Wide Web, it could not have been prepared in the short time with the available resources. The web offers huge opportunities to locate and retrieve information. Search engines like *Google* allow one to mine millions of pages of data – the problem is that one can become almost complacent – "I have so much I must have everything".

In reality, even today's most sophisticated searching misses large amounts of information. Thus academics, working in the biomedical sciences, often give the impression that a MEDLINE search is sufficient to give them the information they require. MEDLINE, excellent though it is, still only covers a percentage of the relevant literature and a proper search should always include additional literature abstract databases such as EMBASE, CAB Abstracts, and PASCAL – to name but three. If one adds the many new curated databases that point to literature that has been searched by hand and/or other techniques, then we see it is increasingly necessary to use a large number of tools to find the materials we require.
But the fact that information is on a site also seems to lull publishers into the false sense of security that users will come to that site of their own will. Several top quality journals, for instance Nature, Science and Cell, have all made statements in the past to the effect that they expect scientists to log onto their sites for relevant material. This attitude is in parallel with an earlier feeling among top publishers that scientists would come to the imprint to look for journal articles – few scientists though can say who publishes which title. In truth, the web has meant we need better and more determined searching services, and OA technology will mean that the net has to be spread even wider and have a variable mesh to get to the detail required (and against this the bench scientist makes it clear that they really require a one-stop-shop with a central "finder" pointing to the relevant titles so that individual articles can be obtained is a high requirement).

One recent advance has been to use the references in a paper to link to other papers of probable interest. The largest service offering this capability is CrossRef (www.crossref.org). CrossRef is a not-for-profit network founded on publisher collaboration, with a mandate to make reference linking throughout the online scholarly literature efficient and reliable. CrossRef is the largest implementer of the Digital Object Identifier System (or DOI - www.doi.org) to date and continues to grow, with 153 publishers using the service (in September 2002 there were some 6,400 journals containing some 5.0 million articles in the database). In the CrossRef system, each DOI is associated with a set of basic metadata and a URL pointer to the full text, so that it uniquely identifies the content item and provides a persistent link to its location on the internet (although the DOI is "outside" the OAI, it conforms to the principles).
Publishers like the CrossRef system because the central service holds no full text but rather effects linkages through the DOIs which are tagged to article metadata supplied by the participating publishers. A researcher clicking on a link (the format of which is determined by publisher preference; for example, a CrossRef button, or "Article" in html) will be connected to a page on the publisher's website showing a full bibliographical citation of the article, and, in

most cases, the abstract as well. The reader can then access the full text article through the appropriate mechanism; subscribers will generally go straight to the text, while others will receive information on access via subscription, document delivery, or pay-per-view. CrossRef costs the researcher nothing; its expenses are covered by nominal charges to the publishers for depositing their metadata, annual membership fees, and fees to publishers of abstracting and indexing databases for accessing CrossRef's bank of DOIs to create links to full text articles.

DOIs are not confined to STM articles. The Foundation stimulates the development of different Application Profiles which stipulate the metadata standards to be used by that application (DOI supports a variety including Dublin Core and OA Harvest). The DOI also allows the group to set the business rules and so allows distinct groups to issue DOIs in their area thus being an ideal candidate to support an international OA regime.
The technology behind the CrossRef service is similar to that being established for some aspects of OA. The publisher must submit a minimal set of metadata about each article, in a defined format (and can submit additional metadata if desired). The DOI/URL will be registered with the International DOI Foundation in the DOI Directory and becomes a fixed reference and pointer to that piece of information.
CrossRef offers a huge advance to the STM community. True, the user can be frustrated when he is refused entry to a linked site because he or his institution does not have a subscription to the located article. To counter this many publishers now allow Pay Per View (ppv), and while the prices per article might seem huge to the average purchaser, this is now an economic rather than technical problem (although many librarians also protest that the user may be taken too easily to a publisher site rather than be allowed to find the material on the local shelves; this can lead to double purchasing unless good access control services identify and warn the user that he/she might be a subscriber and can therefore get the material from the library).

The OA community is starting similar projects. One well established project is The European Mathematical Information Service (EMIS) was established in 1994 and has developed into an information server network of the European Mathematical Society **[3]** . It is based on the voluntary support of several partners from all over the world (Bernd Wegner, and Michael Jost, "EMIS 2001" High Energy Physics Libraries Webzine, issue 6, March 2002, http://library.cern.ch/HEPLW/6/papers/4/.

Citebase Search (http://citebase.eprints.org/) allows researchers to search across free, full-text research literature Eprint archives, with results ranked according to selected criteria (e.g. by *citation impact*), and then to navigate to that literature using citation links and analysis. This service has started to include examples from the Physical, Mathematical, Computer Science, Psychology, Neuroscience, and Biomedical sciences.
CrossRef and other similar linking services allow one to follow a story once it has begun. Traditionally, STM users have used an abstract or citation service for the initial search. Increasingly, many scientists now use *Google*, and other similar services to look for information.

Such a general search system has some attractions and Elsevier has launched Scirus (www.scirus.com) which they claim has a number of advantages over *Google* including the fact that they filter out non-scientific sites, look for peer-reviewed articles from access-controlled databases. For example, if you search on mad cow disease, Scirus finds peer-reviewed articles from ScienceDirect, *Google* finds Web pages only, and also searches deeper than the first two levels of a Web site, thereby revealing much more relevant information. Scirus does not only find Elsevier content and currently includes many of the initiatives currently championed by the OA community, such as the E-Print ArXiv, CogPrints, BioMedCentral the Mathematics Preprint Server and the Chemistry Preprint Server.

## 6. Archiving

Published materials must be available for future generations. It is therefore essential that the STM materials we are discussing in this seminar are "archived" – kept continuously available. The archives must be technically and financially stable, and they must be managed so that the materials are kept in a pristine and non-corruptible state; and they must be accessed by all those who need them: even if this means that some form of payment will be required.
The OAI uses the word archive rather freely and one might be fooled into thinking that materials kept "somewhere in an OAI compatible computer" are sufficiently protected; but they are not. Furthermore, there has been little attempt so far to designate the sort of archives the field requires. The term 'self archiving' is often used in OAI documents to refer to the process whereby individual authors submit their own papers to a server or archive of their choice. Then there are 'institutional archives', where authors submit materials to a server administered by an organisation or scholarly society, or their university or research institute, and there are central discipline-based archives and other speciality archives, presently being developed by the new players in the field (e.g. BioMedCentral).

At the same time, the term is also used to describe collections of information used for a common goal. Eprint or similar archiving technology allows one to collect, with permission, sets of information for further distribution/access. The goals of some of these collections are a little unclear but one important example is the Electronic Research Archive in International Health (ERA), set up by The Lancet. This archive allows medical researchers to deposit papers of special relevance to health issues met in many developing countries. Papers submitted are reviewed before acceptance and are thereafter archived and available online free to all.

Another impetus to archiving comes from the present initiatives to link content. Thus Elsevier, and other publishers such as Springer, are going to great lengths to digitise their back volumes and complete collections. Many publishers are also digitising handbooks and other reference works to add to the electronic content.
But there is archiving and archiving. Obviously a publisher has an interest in maintaining a collection of back issues for back sales. However there is now general agreement that the specialist libraries, and especially the national

libraries, should have a true mandate to archive materials for the future (in many countries, e.g. France, this role is already written into national law). In The Netherlands, the Royal Dutch Library has recently agreed to archive all of Elsevier Science Publishers materials – as a separate initiative from the publisher. They can make these materials available to those permitted to access the library's collections and will ensure that, should Elsevier stop making the titles available on a commercial basis, they can be opened to everyone on a remote basis.

Archiving costs money, and one fear of leaving materials with their commercial owners is that these could go out of business and so the archive could be lost. The same concern should be levelled at the door of non-commercial archiving services, and especially those without a core reason for maintaining them – after all, even the Ginsparg collection moved away from its original home, and who knows if its present home will continue to host it without income "for ever". Many new initiatives have lost their contents as they closed their doors when initial grants were cut. Many university treasurers are less inclined to spend money on storing the reports and papers of faculty who have moved on than the scientist himself might think. Some "new homes" will be wary of taking responsibility for materials produced under another regime. And some reports might get "lost" in an archive that is more devoted to other activities and subjects. Thus it seems a little naïve to expect any institution to allow "self-archiving" without some firm intentions and rules.

Even National Libraries do not have infinite funds but these institutions seem the best to spearhead the archiving initiative. Before they do so however they will surely want to ensure that what is archived is essential and good quality. Materials should not be duplicated, nor should non-definitive materials necessarily be archived when a final version is stored elsewhere. The amount of material involved is so huge that efficiency will be the order of the day.
Recently the International Committee on Science and Technology Information (www.icsti.org) has started to try to increase awareness amongst the scientific, publishing, library, archive and data communities of the need for action to preserve the digital records of science.

ICSTI feels that a comprehensive scientific digital archive is likely to involve a complex network resulting from discipline-specific, institutional, national and international initiatives. Standards therefore need to be chosen and adhered to (an important model appears to be the Open Archives Information System (OAIS) Reference Model, which will be an ISO standard, http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html). ICSTI certainly considers that further work is required to define archiving policies, to be clear about where responsibilities lie and to ensure that a supported, funded and sustainable infrastructure is put in place which can stand the test of time.
Even so, the ICSTI Advocacy Statement on this issue raises more questions than answers:
· Is it merely the content or also its form and presentation which need archiving?
· How is authenticity to be guaranteed?
· Which version(s) of should be preserved?
· Are the links, which are so important in many electronic information services to be maintained and, if so, how?

· How are copyright and other IPR issues to be addressed in ways which enable repositories to maintain perpetual archives and provide appropriate access to users?

The ICSTI statement mentions other issues, for instance, can one harvest material from archives where the copyright might rest with the author? Or, will the data protection and database harmonisation laws prevent us using the technology we have to the full? (The Copyright and IPR legal and practical environments have received a great deal of attention since the coming of the internet. The European Commission are presently addressing these in a series of directives, seminars and working parties. Two reports, relevant to STM, on IPR and Bioinformatics, and IPR and the Internet, can be found at http://europa.eu.int/comm/research/era/ipr_en.html.)

One way in which networking is leading to archives is described in Jeffrey Young's article 'Superarchives' Could Hold All Scholarly Output, http://chronicle.com/free/v48/i43/43a02901.htm. Young points out that many institutes are exploring bundling their own materials so that they can be exchanged and worked on in a "free environment". The article does not indicate what quality control gateways these archives might establish, although it does suggest that "editors" might be appointed, and tries to keep the content to the sort of material that would/will be published anyway. The initiative clearly shows that a number of top institutes see the archive as a "calling card/advertisement" for their excellence. Again it appears that we might get a world of "rich" and "poor" centres with the rich offering a range of tools and materials in the form of interactive archives while the rest struggle to offer what they have in a less efficient manner.

One important factor needs to be further investigated. It is common for people to look at new versions of old services as being "more or less the same". Thus, presently, the STM article is often considered as being "text on a screen" rather than on paper. But modern technology offers a lot more and soon STM "articles" will be supported by other materials and services such as video links, audio files, illustrations, tables, and links to software programmes where data can be manipulated and visualised. Many of these developments are being examined in the emerging digital libraries field (for instance go to the European based group – DELOS – at www.delos-noe.iei.pi.cnr.it/). These developments have implications for archiving, not lest the need to make available in perpetuity the readers for these files, but they also may create more barriers to access for those who don't have the requisite bandwidth and equipment.

## 7. Subscriptions, Licences and other barriers to free access.

The ultimate goal of scientists searching the STM literature is to find full text article they require and, as said earlier, if one asks most academics in biomedicine, they seem satisfied to have covered the literature with a MEDLINE search.

This is dangerous. Estimates differ but there are probably more than 20,000 journals of some relevance to the scientist and MEDLINE only abstracts some

4,000 of them. Even adding the other abstracting services together leaves a number of titles more or less uncovered. New services are coming to the arket but they too seem to concentrate upon the 'best' or the 'top' articles (e.g. the Faculty of 1000 project www.facultyof1000.com).

OA could therefore offer access to much of this "hidden, or less visible, science". Some of it will be on publishers servers as most organisations now offer their current, and some back, content, online; either on their own web site or on an aggregator's such as Ingenta (www.ingenta.com) or Highwire (www.highwire.org). Therefore, we should be able to get to the article if the accompanying metadata allows one to locate it, or the correct address/link can be found (and a service such as CrossRef will naturally make this easier).
A major problem for the user however, is that this content is protected from casual access. Publishers have invested heavily in building their content lists and so therefore insist on selling this, usually as far as journals are concerned, by subscription or license, rather than by selling the individual article. There are statements that these licences will one day become so flexible that "anyone" can obtain "access to anything" but this is not the general case as yet.
No-one can doubt that the commercial success of the present commercial STM publishing industry lies in the subscription model for journals: even scholarly societies have been known to launch new journals just to add to their income. There is also little doubt that the success has led to what many buyers see as exploitation and the ever increasing prices have led, in the past decade or more, to ever falling subscription lists.

The result is that most publishers in the STM sector now rely upon a small core of R&D orientated libraries for their sales. Smaller institutes, departmental libraries and individuals have cancelled many of their holdings so that few libraries can offer the service their librarians wish. The publisher's income has also become more and more vulnerable as it now rests upon a small number of paying centres.
The "electronic revolution" now offers publishers a new opportunity. Electronic publishing costs are "up front". Once a publisher has established the infrastructure, access control and content in digital form, adding new users to the community is cheap. The major publishers have therefore been making huge efforts to switch to licensing, e.g. by site or consortia. By converting their major "individual title subscriptions" to licenses for "collections of content", they have managed to maintain their "pre-paid" financing but have also offered the user more. Another development has been the increased movement towards co-operative or consortium purchasing so that smaller institutions can join forces with others and buy a consortium license. This has brought many smaller institutions who had previously cancelled materials back into the picture.
At first sight, this is succeeding. Elsevier Science reports that they are earning some 65% of their journal income from site and other journal access licences (starting from zero some 4 years ago). John Wiley and Sons report similar successes. These and other publishers are seemingly delighted with the results they have achieved, and hope and claim that the market is also pleased. Usage is up, those with a licence download huge numbers of articles (but what is usage - if it costs nothing most scientists will download an article for a look later and there is no guarantee that a downloaded article is actually read) and far more readers can access the journal articles they require. Thus, why change the system? If you want an article join someone who has a site licence.

The immediate counter argument is that even though many in the scientific sector might be affiliated to an institute who can be affiliated to a consortium or site licence, there are still many that are not. How do these get access?
This issue is also being looked at. As it costs so little to add access for new customers, the major publishers are looking at new license forms – and even at whether or not differential pricing for different countries might not be possible. There are even indications that the site licence system could be expanded to libraries and public libraries.

Superficially, this might satisfy everyone's need. However, there are still many people who question whether licences can satisfy the market. For instance, the publishers are tending towards offering larger and larger collections. However, the market is asking for specialist collections of articles selected along defined/described lines. Furthermore, offering a "catch all" of materials under a supposedly "quality blanket" (i.e. all our material is good) will just stoke the fires of user discontent – publishers have good and bad titles in their collections and the market knows that all too well. Leo Waaijers of Delft Technical University Library put this into perspective by reporting the point that: "Consortia are only good for publishers, not for libraries. We now have to buy 1100 titles from Elsevier while we really only need 68" (see, "A new role for the Scientific Publisher in the electronic age" www.iospress.nl). Furthermore many librarians feel that they need something more tangible at the end of a contract, there is all too often no paper version and so archiving and back filing becomes an issue.
The size issue is of course a factor in the real problem: the cost of the information. More and more organisations are admitting that they cannot afford the information they require. Furthermore, some complain that the pressure to buy into a license means they have no funds left for other materials. If a license covering a substantial, but not total, amount of information costs so much that there is no money left for other materials from other publishers – what goes? And who decides these policies? A multi-disciplinary library might appear to be the best partner for a publisher to negotiate a license with, but there is also evidence that many specialist libraries feel they need to look for new model. The bottom line seems increasingly to be that it is no use including materials in a license agreement that no one requires if they then have to forgo materials from other sources that are essential.

Recent research carried out by, and reported during, the Ingenta Institute meeting (to be published later this year – see www.ingenta.com ) also indicated problems for the immediate future. The "Big Deal" - the bundling of all titles from one publisher within one licence agreement either to a single library or increasingly a consortia of libraries - made sense a few years ago when libraries were able to attract new funds to purchase such bundles. However, now, such start-up funds have all but ceased and, 3-5 years later, some of the consortia are facing renegotiations of the original deals and finding they cannot afford it.
Perhaps this all boils down to a new version of the 'frustration gap' in scholarly communication – static library budgets trying to buy information which is growing at 5-8% per annum, excluding inflation. However the problem is real: the cost and size of the information on offer is rising while library budgets are stable (at best). The "Big Deal" has merely postponed the evil decision day by the once-and-for-all attraction of exceptional government funding for the start of the licensing deals. The consensus from the Ingenta Institute research was that these deals would break down in future. Furthermore, there is probably going to

be little extra money available for the funding of any major new investments in the STM information market. In any case, the "license solution" has not stemmed all the cries of customer discontent in the STM market.

## 8. Costs

There is no such thing as a free lunch and someone, somewhere, will be paying – whether that is for a journal or an "open archive". Even the much applauded academic library "Inter Library Loan" system appeared to be "free" but actually cost the participating libraries a lot of money in hidden administration and labour costs.

There are numerous estimates as to what it costs to process (publish) a journal article. The new company BioMedCentral is applying (semi) commercial rules to OA policies and are presently charging an Article Processing Fee of $500. This is thought by many to be too low but it at least an indication of what is required for a refereed paper to be processed and released under OA guidelines.
Archiving will also be expensive and one might just as well look at the total costs of the STM publishing world to get an idea of the sort of budgets needed. And with journal growth, i.e. manuscript flow, increasing at around three percent per annum, the costs will continue to grow. Who will pay for this?

A core argument today, is that the "state" pays – through financing the libraries. This is felt by many in academia to be unfair as the "state" paid for the research in the first place and now has to pay to get at its own research.
On the face of it this makes sense. However, "publishing" costs money. If one looks at an R&D grant there is rarely a budget line for "publishing" – the assumption is that the work will result in a paper that will be published at someone else's financial risk (the publisher).
In simple terms the library budgets are therefore too small. But turning to the "page charge model" will also mean that money has to be found - presumably moving grants from the library to the research department so they could publish their materials instead of buying information. Is this what the academic world wants? Will librarians allow this sort of transfer of funds (and influence and power?).

This might appear too simplistic, but is indeed the model followed by the funding of major technical and information projects like the Human Genome Project (and other genetic information databanks). Basically the granting bodies fund the institutions, in the USA, Europe and Japan, to make their data available for free.
In simple terms, either the centre is paid to give their database(s) away for free, or the (academic) is paid to buy the data from the centre.
The OAI is being funded from central public sources and many new public sector initiatives are adhering to their guidelines. In Europe, E-BioSci (www.E-biosci.org) will be OA compliant in its plans to offer a raft of new information storage and retrieval services in the biomedical arena. When operative, the project will allow the user to navigate between bibliographic and other factual and sequence databases and to the full text of the relevant article. It will also establish a new digital publishing service where authors can deposit their

research article on the web and so, in the long term, build a European based research infrastructure.

Part of E-BioSci will therefore challenge the existing status quo in STM publishing. Another such challenger is FIGARO (www.figaro-europe.net ) a collection of European universities and publishers which aims to create a European network of institutions providing e-publishing support to the European academic community. FIGARO will investigate new business models for scholarly publishing and will stimulate open access to the publications produced and distributed with its infrastructure, making scholarly publishing faster, cheaper and simpler.

The inclusion of a "business model" in Figaro is a welcome note. However, are other projects being forced to look further than their present grant? For instance, some of what E- BioSci intends to do, linking from nucleotide sequence databases to the scientific literature, is not difficult and has actually been done before in the EC funded project ADLIB - Advanced Database Linking in Biotechnology. This "proof of concept" project was a public/private co-operative which solved these problems and proved the technology as early as 1998. However, the teams failed to find a way of making money from the service and so abandoned their plans despite the obvious interest in the service as such. ADLIB did not apply for more public money but might the European Commission have to continually fund E-BioSci? And is this support to be open-ended and non-capped? If so, it might soon swallow a large amount of the EU's R&D funds unless new budget lines are developed to sustain such central (as that is what this would become) services.

For if OA is to make any long lasting changes to the way information is handled, it must be sustainable. There are many "rich" organisations supporting the initial OAI projects and archives but they alone cannot handle the volume of material that one will encounter if the OA dream becomes reality.
This ultimately means long term financing, which is free from political influence. Grants are notoriously poor at launching novel industries. Looking through the EC's records on funding novel projects, one finds few that have actually gone on to stand on their own economic feet. Nevertheless, the European Commission is increasingly clear in stating that, eventually, the projects they support have to be self-sufficient. This is not just a political statement to satisfy free-market supporters - the Commission is unable to guarantee funding over a period longer than one Framework Programme and there is no guarantee that projects funded in one framework programme will even be allowed to apply for money in the next (i.e. priorities and regulations can change dramatically **[4]** ). This means that, in the context of EU funding, four years is the maximum time a project can be guaranteed funding. Many commercial publishers would be delighted if they could get their products breaking even, let alone making a profit which would enable them to survive on their own two feet, in that time. Information needs continuity and continuity needs funds; and public funding is not really in a position to give that sort of life-long guarantee.

Nevertheless, OA, like many other innovative activities, can be hugely helped by initial public funding. Afterwards however, it will probably benefit more from applying good business acumen, although there are not too many new options other than those used by the publishing world today.

William Y. Arms offers an independent list of opportunities in his "Economic Models for Open Access Publishing" (http://www.cisp.org/imp/march_2000/03_00arms.htm) . However he finds no new solutions offering four models of funding for OA projects, namely:

· Restricted access: use-based payment,
· Restricted access: subscriptions,
· Open access: advertising, and
· Open access: external funding.

At the moment, only the last seems to fit both the political and economic needs of the OA community. The real solution will probably be a combination of public and commercial initiatives – as we have today. At one end of the scale this might involve advertising. Several new people in the OA community hope that it might provide up to 20% of their income, which with Article Processing charges might give sufficient income to run OA journals. Others are sceptical, pointing out that advertising has consistently failed to live up to expectations in this medium. So subscription models might well be seen again – however much they are disliked.

One solution might actually come from an increase in transparency surrounding the financing of running such services. Arms, again, points out that at Cornell University, the aggregate of funds that individual departments spend on their web sites is greater than the library's total budget for acquisitions. Running an internet node at a university costs far more than the users might believe. There has already been a huge shift in expenditure within institutions and so OA might become an "infrastructure expense" that is covered by the organisation as other infrastructures are today. Still, one is presently looking at opinions and hopes – there is certainly no concrete plan for the long term funding of OA.

## 9. The Developing World

It has often been said that the evolution of electronic publishing will allow the developing world to catch up with the developed in terms of accessing and using STM, and other, information. No longer will poor countries have to purchase large libraries, which they cannot afford, but they will be able to access central collections at will. "Divides", or barriers to information flow, such as the North/South, South/South, and South/North can all benefit; especially for the latter two where the new technologies allow scientists to broadcast their work to a huge audience at little cost (for instance, see www.ariadne.ac.uk/issue30/oai-chan/).

OA offers the Developing World the option of making their materials available for all to access and use. In theory, academics in poorer countries can take advantage of servers anywhere in the world offering OAI services, without the need to set up their own independent servers. A current major obstacle is the lack of awareness of the availability of the different mechanisms for distributing and accessing research documents and, as these are usually internet dependent, this is exacerbated by the lack of telecommunications infrastructure in the developing world. However, there are major international and local efforts to invest in the infrastructure and there is growing optimism that with time, this problem of a 'digital divide' will be resolved.

The best known of the initiatives which specifically note that they might help the Developing World is the Budapest Open Access Initiative (BOAI http://www.soros.org/openaccess/read.shtml). Launched at the end of 2001, the initiative is based upon the belief that science is … "published without reward to the author for the sake of inquiry and knowledge" (and one should also add, visibility). Taking this premise to its logical conclusion it follows that, … "if this information could be freely accessible this would benefit the world even further… Removing access barriers to this literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge".
The BOAI, while confronting the present commercial publishing world about prices, do admit that STM, peer-reviewed, literature costs money to produce. The initiative's argument is that the profits made are too high and so the initiators are supporting efforts to break the present mould and replace it with a new infrastructure. The group therefore recommends that:
" … To achieve open access to scholarly journal literature, we recommend two complementary strategies.

1. Self-Archiving: First, scholars need the tools and assistance to deposit their refereed journal articles in open electronic archives, a practice commonly called, self-archiving. When these archives conform to standards created by the Open Archives Initiative, then search engines and other tools can treat the separate archives as one. Users then need not know which archives exist or where they are located in order to find and make use of their contents.

2. Open-access Journals: Second, scholars need the means to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access. Because journal articles should be disseminated as widely as possible, these new journals will no longer invoke copyright to restrict access to and use of the material they publish. Instead they will use copyright and other tools to ensure permanent open access to all the articles they publish. Because price is a barrier to access, these new journals will not charge subscription or access fees, and will turn to other methods for covering their expenses. There are many alternative sources of funds for this purpose, including the foundations and governments that fund research, the universities and laboratories that employ researchers, endowments set up by discipline or institution, friends of the cause of open access, profits from the sale of add-ons to the basic texts, funds freed up by the demise or cancellation of journals charging traditional subscription or access fees, or even contributions from the researchers themselves. There is no need to favour one of these solutions over the others for all disciplines or nations, and no need to stop looking for other, creative alternatives."

It is clear that these ideals are close to others in the field but the finding of real solutions is still rather general and vague. It also almost seems naïve to list possible funding sources without indicating the way in which these might be involved. It will surely require more than a few "statements" to organise the alternative to the present STM information infrastructure with a large-scale change of funding as well as changes to established habits and even scientific

measurements (citation measuring is to a great degree tied to the ranking of titles as we have them today).

Advocates for this change are naturally hoping that the BOAI will succeed due to being financed by The Open Society Institute, the foundation network founded by philanthropist George Soros. It is clear however, even to these well-funded people, that the initiative will need other organisations to lend effort and resources as well as a high degree of bureaucratic planning and administration to change from what we have to what they want.
OA can probably succeed by following the same path that the internet has: every player will have to add a little to the infrastructure so that the sum of all is equal to a new environment.

## 10. Where is it all going?

There is little doubt that OA and the OAI will open up new avenues for the storage and retrieval of STM information. The technology will soon exist for users to carry out a federated search over a number of disseminated centres and archives. Individual and clusters of like-minded institutes will put their files online so the content can be aggregated, mined, or linked to added-value services. Digital libraries will establish collections of information which has been validated and expanded to include background and supportive material; and scientific articles will evolve with criticism and comment from the pre-print environment.

But will this replace the present STM information system, and in particular the scientific journal? After all, the publishing industry is established, well funded, and has a strategic lead in the actual use of new technologies. The publishers also have a proven and efficient infrastructure with many millions of pages of content which itself forms the basis for new services and advances.

The immediate conclusion must therefore be no. OA is still a set of evolving technologies. It does not (yet) possess a "ready-to-go" infrastructure. There are doubts that some of the harvesting aspects of the OAI will actually do what is promised, and the new initiatives are presently too small and are moving too slowly to seriously affect the present system; and they lack the financial basis to persuade people to trust and commit to them for the long term. Furthermore, the OAI is still very much a "club" of like minded enthusiasts. While they are certainly addressing utterly important issues, the debate is still muted and the discussions on these issues only involve a minority of scientists – the wider debate on funding information has hardly begun.

At the same time, if anything does challenge today's publishing status quo, it is the cost of today's journals. To put it bluntly, the market now perceives "journals" as being too expensive. Critics of the commercial publishing world claim that publishers have controlled the STM market with ruthless efficiency. Annual journal prices have been raised above inflation, extra volumes were added to existing packages without any chance of the purchaser being able to remonstrate/refuse, and new journals were added while it was known that the market was already well endowed with similar titles. The cost of journals rose

and rose and today, many feel it has reached a breaking point - there is just not enough money available to cover the prices charged by the publishing industry. Without wishing to enter into a debate on the rights and wrongs of the commercial publishing world, it is obviously important to say that the publishers do not generate the articles they publish themselves. Scientists have to publish. A scientific article is the recognised end-point of a research project and the success of the science will very much relate to the success of the article. The cost of journals is, to a great degree, because they include a lot of material. Unfortunately however, most of the articles only appeal to a few other scientists and the ratio of articles which generate a large amount of interest, and therefore readers, is small compared to those that are relatively esoteric.

The esoteric articles still cost the same amount of money to process as do the interesting articles. Therefore the market turned to the subscription model where one accepts that few of the articles "bought" will be of real interest. This model has generated the large sums of money publishers make and use to continue publishing. Without these sums of money, much less material would be handled, but the price of the essential articles might be smaller.

This debate has raged for decades and will continue. But perhaps OA can play a relevant role in the overall solution by, for instance, allowing some forms of information to be available "free". Thus we might see the role of the journal changing.

For the moment however, those advocating change seem to be trying to reproduce the journal – but make it cheaper. The Public Library of Science (PLoS www.plos.org) initiative started as a protest against the cost of traditional journals and the fact that few are available for free on the PubMedCentral server in the USA. The initiators first called for an author's boycott of those journals and over 30,000 scientists pledged to do so. However, to-date, there has been little effect. Manuscript flow to the journals in question appears to be more or less the same as before PLoS made its call. The PLoS group admit they do not have any concrete information as to whether their initiative is working but do feel they will have to start a competitive service if they are to "win". Thus they report (in a letter to the author, September 2002) that:
"(we are) on the verge of an announcement of strong financial backing for our efforts to launch our own journals. We anticipate that they will begin accepting articles in the first half of next year and begin publishing next summer."
The PLoS also point out that articles submitted to BioMed Central (an open access publisher) are constantly increasing, and take this as a measure of success in their campaign.

BioMedCentral is funding its publishing business through Article Processing Fees (or page charges as they are traditionally called). BioMedCentral is hoping that $500 per article will be sufficient for them to finance their service but admit that this must be checked in practice. BioMedCentral has a well-defined charter that goes a long way to satisfying PLoS demands and so should satisfy the ethical points being made (this charter includes the fact that authors hold the copyright of their article and that the articles are mounted on PubMedCentral and other archiving services). BioMedCentral is particularly determined to ensure that articles stored with them are kept for the future – an indication of their

commitment to new technologies but also to the need for experimental services to ensure that what they publish is retrievable for time to come.

In truth, BioMedCentral is still a very small player. Time will tell whether the PLoS group can start their challenge but even so many question whether they will dramatically alter the submission patterns to major journals. There are more articles appearing each year and some will undoubtedly go to new projects but it is a brave scientist who throws away the chance of a top citation in a leading title. Furthermore, despite their worries as to the cost of journals, university administrators and grant funders are still keen to see the fruits of their funding and effort published in the top titles. Not being cited, or being placed in a journal that is not abstracted in MEDLINE or one of the other big abstract services such as PASCAL or EMBASE, is "death" to the information; and in some cases it might mean the loss of grants as some funding bodies still use the citation impact figure of a scientist's output as a measure of success.

So, if BioMedCentral starts to grow exponentially, and PLoS do manage to start an alternative publishing service, then we might see the start to a change in publishing. However, the signs are that this will take many years before it can significantly challenge the present infrastructure.
Above all, scientists will need convincing that their work is not going to be lost. The present journals do lend a feeling of permanence that a server tucked away on a campus does not. Thus if this issue can be handled, a better argument for OA to challenge the present status quo might be the simple one of visibility. Steve Lawrence's contribution to the Nature debate on e-access (Free online availability substantially increases a paper's impact http://www.nature.com/nature/debates/e-access/Articles/) shows clearly that authors in computing science can expect higher citation figures if their articles are freely available online. This will almost certainly be mirrored in other disciplines. This then can lure the author, and challenge the publisher to make their articles more visible and accessible. Again, this might not mean that scientists will turn away from journals, but rather that journals find some way of making their articles more visible.

Overall, it is also important to remember that few if any publishers oppose the idea of OA. The opposite is probably true and many are enthusiastic supporters of standards, and linking, and common searching engines. After all, commercial publishers have, after a slow start, invested huge sums in the internet revolution and it is in their interest to see the service succeed. Commercially, it is hardly surprising that the larger publishers are using their present subscription portfolios to leverage income from the new electronic world and so one might actually expect to see the commercial world use the OA initiatives more rather than less.

The supporters of Eprints and self-archiving hope that a self-archive will contain pre-prints and post-prints and that eventually the post-print will be free. This seems unlikely – publishing costs money and in a market economy this has to be earned – but perhaps the two environments can complement each better than many think. In fact it makes far more sense for the commercial and academic sectors to come together to develop some sustainable future for the STM information world than to compete with each other when funds are so tight.

Therefore, instead of trying to mirror the journal, a better way forward might be to use the OAI to generate new ways of information dissemination. One no longer has to go to a library for information – one can use the internet to go to many different sites at once. The journal is also no longer tied to library shelves but can be used as a dynamic source of information. Furthermore, the journal is no longer limited by the limitations of paper which is certainly no longer the best medium for presenting and handling modern data. Much information has to be analysed (and therefore refereed) using specialist computer programmes. "Papers" often publish the Final Results, ignoring the data that the authors analysed on their way to their conclusions. There have been rumblings in many fields, for instance crystallography and clinical trails, that papers should be accompanied by data that can actually be analysed and many scientists would surely welcome the opportunity to mount supportive illustrations and other data behind their paper.

To-date, the journal article is more often than not the end-point, the proof that the work was executed. Increasingly, that measurement of success will also come from the analysis of an author's data deposited in databanks. The journal article will then become just part of an information story and publishers might then find that they are offering just one part of an interconnected whole. This might actually reduce the amount of information requiring formal "journal publishing" meaning that a shift occurs in the whole market. Publishers might handle just the "top" materials that require in depth refereeing, or offer reviews which tie a field/discipline/sub-discipline together. They might further develop, and charge for, the complete information chain, ensuring that scientific information is part of the research process rather than being the "end report". OA might, in parallel, offer a way to find and obtain information before it reaches this "set-in-stone" phase. Some materials will probably therefore not reach a journal but will reside on pre-print servers or in databanks.

Finally, a pause for thought. Like it or not, there are few things in our modern world that do not have a price tag. And there is always someone paying - even for a free product. Ironically, many things can be free while they are unsuccessful, and small and therefore cheap, but quite often successful *free* services grow to the point that they cost too much to be free! Therefore there is every chance that a totally successful OA based scientific information service would eventually have to be charged for. One portent for the future is perhaps the Journal of High Energy Physics (JHEP http://www.iop.org/journals/jhep), an OAI compliant publication that has been offered free-of-charge since its start in 1997. The electronic journal was published by the Institute of Physics Publishing while submission and peer review were the responsibility of the International School for Advanced Studies (SISSA) in Trieste, Italy. Until now, the journal has been free due to contributions from the parent bodies but, as it has become more important and grown in size, this route of financing is no longer viable. Thus, as of January 2003, the journal will be available to institutions for an annual subscription price only: although the back archive will remain free.

**[1]** The searching software SRS, which is in fairly ubiquitous use in the bioinformatics world, was developed by scientists at EMBL, with help from the EMBnet group. It was later privatised by LION Bioscience (www.Lion.com) but, at the time of privatisation, was a mainstay of the academic bioinformatics infrastructure. There were many who were therefore unhappy at this turn of events but Lion gave an undertaking that they would continue to offer academic and related centres a free license. This has worked well, and, at the time of writing, (Release 7) the software remains the same to ensure that all SRS sites remain compatible. The commercial license customers receive a large amount of support and related expertise.

**[2]** The following indicate how publishers across the STM field are moving towards allowing self-archiving:
Nature states: "Nature does not wish to hinder communication between scientists... Neither conferences nor preprint servers constitute prior publication."
Elsevier allows: "As an [Elsevier] author you [have the] right to retain a preprint version of the article on a public electronic server such as the World Wide Web."
The American Physical Society 's copyright statement encourages self-archiving:
" The author(s) shall have the following rights... The right to post and update the Article on e-print servers as long as files prepared and/or formatted by APS or its vendors are not used for that purpose. Any such posting made or updated after the acceptance of the Article for publication shall include a link to the online abstract in the APS journal or to the entry page of the journal."
The Association of Learned and Professional Society Publishers' (ALPSP) model license states:
" You... retain the right to use your own article (provided you acknowledge the published original in standard bibliographic citation form) in the following ways as long as you do not sell it [or give it away] in ways that would conflict directly with our commercial business interests. You are free to use your article... mounted on your own or your institution's website; [posted to free public servers of preprints and/or articles in your subject area]..."

**[3]** A key feature of the service is the Electronic Library of EMIS (ELibM) which presents a collection of freely accessible electronic publications. The aim is to be as comprehensive as possible although some growth is required. The service has developed its own refereeing service, based on the criteria used by traditional journals, to ensure good quality
EMIS recognises that some publishers will be unwilling to give away their income and so most of the journals on offer are published at a low-budget level, and hence the risk of loosing subscribers to the print version due to the free electronic offer currently is considered low. Some of them give the electronic offer with a certain delay to EMIS such that the earlier availability will be considered as an advantage of the print version.
EMIS also links to databases including MATH - the online version of Zentralblatt MATH, MATHDI - the online version of a similar service for education in mathematics, and MPRESS - a global pre-print index, and a database on geometric objects.
The service has obviously been developed on feedback from the membership. It has been supported by a number of grants but relies heavily upon volunteers. Furthermore, the full service has to be subscribed to.

The authors make the point that EMIS could not survive on the current level without the big group of its supporters, who serve as volunteers for maintaining and installing electronic journals, caring about submissions and transfer of content and keeping the mirrors running. Without these activities EMIS would not have been possible. This indeed shows that a viable service can be maintained with the collaboration of several volunteers, but it does not prove, as they appear to claim, that the service could grow to the size and completeness needed to cover mathematics globally. This does not also mean that such a service has to be commercial – although they do charge a subscription fee so what is commercial – but it probably means "is there a profit element involved"?

**[4]** The protein structure database SwissProt, recognised as an essential cornerstone of modern bioinformatics, was crucially supported by a series of EC grants. Following the adoption of a new Framework Programme the EC funding was not renewed. The producers then had to adopt a rushed policy of charging commercial users a fee for the use of the database thus turning a free product into a commercial one. The moves antagonised many and caused severe concern in the user population. SwissProt is now exploited by the Swiss Biotechnology Institute being funded by a combination of commercial fees and grants.