Establishing an e-print repository: economic issues and practicalities

Alan Hopkinson

Middlesex University

Abstract

The Sherpa project of UK JISC is investigating the issues of setting up e-print repositories. Many universities in the UK are building these. After an introduction to the state of the art in the UK and elsewhere, this paper outlines the reasons behind the setting up of an e-print repository at Middlesex University and describes the efforts made so far. It then concludes with the

1. Introduction

What is an e-print repository? It is a kind of archive of printed material converted into electronic form or material that was born digital. Today there are around the world many sets of academic publications consisting of files stored on servers. Many academic institutions encourage their researchers to place material which they have produced on websites. These would usually be open to anyone to view. Academics place their CVs on websites with links to lists of publications and link from these lists to the actual text where available. Some university faculties have departmental websites where they organise their research output for the world to see. Some institutions positively encourage, even coerce, their academics to make available their materials on servers within VLEs (Virtual Learning Environments) but these are accessible only by their very nature to students on particular courses. Thus files of learning materials may be found in many different locations with varying access rules. An e-print repository brings together all the material of a particular provenance, for example an institutional repository brings together all the papers produced by the staff and students of a university into one place. From the other locations mentioned above links can be made to the files in this one place. Another feature of the e-print repository is that it is part of the open access movement set up to remove impediments to academics finding the articles they need. In the days of printed journals, a library held a journal or it did not. If it did not hold it the only way the librarian could obtain an article from a journal was by inter library loan or by buying a journal issue if it was not out of print. Today in the electronic world it is not so clear cut. Moreover journals are becoming more expensive but publishing on the web is becoming cheaper. Libraries may not hold an article but it may be available somewhere on the World Wide Web, but we may not know where to find it. The only barrier to academics publishing on the web is that they fear that they would lose the prestige and the authentication of a refereed journal.

Why are these called repositories? Another word that could be used is *archive* but this has associations with the historical field and the proponents of the e-print repositories tend to be from the scientific disciplines. Therefore the word repository has come to be used for these.

2. e-Print repositories

It is one thing to make available material on an e-print repository, it is another to ensure that the material is found by those who are searching. To facilitate this, most e-print repositories are set up using a software protocol called the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). This is a method of ensuring that searches can be made across the indexes of different repositories Once a repository has been registered as conforming to the standard, harvesting services will automatically search the metadata

Another feature of repositories is that they tend to be set up by libraries rather than other faculties or departments in universities. Libraries are well placed to set up repositories: they have the expertise to catalogue and classify material. They also have an interest in methods of solving the problem of ever increasing journal subscriptions. They are also responsible in one way or another for making material available to students and this is one way of making one's own institutions output available.

There are a number of reasons for setting up repositories on an institutional basis. Lecturers in an institution may wish to offer their papers unrestrictedly to students since they would traditionally have told them to read their journal articles and offered a photocopy to the library to facilitate this: this is a way of doing the same thing in the electronic era. An institution may like to know what it has produced for research assessment exercises. If all its output is in one place it becomes much easier to evaluate this. There may be problems of copyright in mounting this material in a publicly accessible repository. This has meant that some academics have been reluctant to use repositories, feeling they might not be able to publish an article in a refereed journal with the academic kudos that that brings. Some publishers have agreed that their authors may place on the repository a pre-refereed article or articles in other stages of the publication chain. Other publishers have become more generous as we will see later, through economic and political persuasion.

Although libraries have become the enthusiasts in this area the project was begun and is still 'managed' by an academic, Professor Steven Harnad, Professor of Cognitive Science, University of Southampton. It is at Southampton that GNU e-Prints was developed, one of the freely available software packages that may be used to create a repository[i].

The origin of the movement was probably at Los Alamos National Laboratory where in 1991 a repository of papers on high energy physics was established. Though it was set up as a means of circumventing traditional publishing media, it became the primary means of communicating ongoing research information in high energy particle theory[ii]. This has since extended into other scientific areas: ArXiv as it is now known is an e-print service in the fields of physics, mathematics, non-linear science, computer science, and quantitative biology. now owned, operated and funded by Cornell University, a private not-for-profit educational institution. ArXiv is also partially funded by the National Science Foundation in collaboration with local support .

Another champion, again institutional, is the Open Society Foundation based in Budapest founded by founded by George Soros and originally serving the Eastern bloc before it spread its influence more widely into developing countries. This has set up the Budapest Open Access Initiative. Furthermore software and advice is available on their website[iii]. They work closely with Southampton University.

There is a project being funded by the UK Joint Information Services Committee called Sherpa[iv] which aims to investigate issues to do with the future of scholarly communication and publishing. In particular, it is initiating the development of openly accessible institutional digital repositories of research output in a number of research universities. These so-called 'e-print archives' will contain papers by researchers from the participating institutions. The project is investigating the Intellectual Property Rights, quality control and other key management issues associated with making the research literature freely available to the research community. It will also investigate technical questions, including interoperability between repositories and digital preservation of e-prints. The University of Glasgow is a member of Sherpa. Their website is displayed on the previous page.

3. Setting up an e-print repository

One would like to think that the provision of any new service is done on a rational basis and mostly they are. The reasons behind the provision of an e-print repository seemed quite reasonable, both at the time and retrospectively. It is often difficult to pinpoint the exact time an idea comes into one's head but in this case it is not so difficult. I lectured on electronic journals at a training course for CPD25, Continual Professional Development for London Higher Education Institutions and another speaker was Stephen Pinfield of Nottingham University who is Director of the SHERPA project. Stephen was certain that it should be the library that should provide this service rather than say the university press or the university registry or anyone else for that matter. So I resolved that this was something we should try and do.

I was just wondering about how we should begin to implement this which would surely require a great deal of effort because it meant liasing with academic

staff with whom I do not have very much contact. Then a new member of staff joined our library systems team who was in the second year of a part-time librarianship master's course and wanted to do a project on e-print repositories. He was mostly concerned with evaluating the repository but to do so had to set one up. So that is how we made the decision to set up an e-print repository.

Firstly he had to decide which software to use. There are two main contenders, both available free of charge, and we chose the GNU. We applied for money for a computer from a fund which did not offer hardware as it turned out so we were unsuccessful. They said we could apply for funds to produce publicity to students or staff but not for hardware. Nevertheless we found an old computer we could use as we did not need a big computer for this kind of server, at least not in its pilot stage, and so we were ready to start work. Our colleague had to install LINUX as GNU works only under UNIX. Then he set up the software which is available free of charge from the internet. Then he had to set up sample data, so he used a few articles which people in the library had written and set up a pilot.

There in fact the story ends for the moment. Our colleague left but he was only responsible for the technical aspect. The next part will be more difficult, persuading staff to use it. In Middlesex University the School of Computing have in the past been very strongly involved in the user aspect of digital libraries and have done some work partly to support research and partly as consultancies for people wanting digital libraries. They could probably set up their own repository on their own server and of course it would not be impossible for some of the other Schools to do this, but it would be a pity not to have a general Middlesex University server. This responsibility for seeing that the server is used has been taken on by Sue Hurley who is database librarian based at the campus where most of the Business School is situated. We decided to pilot our server there for a number of reasons. Library staff at that campus were more interested. We felt that business studies are an area where placing articles on a repository would be attractive. The pilot has only just started and in the vacation we found that staff were on vacation and could not contribute. In the term they will be too busy. The few staff approached seemed a little reluctant to commit themselves to placing their articles in the repository, perhaps because they were not sure of the legalities.

If our repository is to get off the ground we will have to make it accessible to the outside world: it is not currently. We will also have to build indexes (see the University of Glasgow website for the kinds of index they have set up). We would ultimately have to set conditions for contributed articles but this is not work that we would expect to come from the library.

4. Other experience

An unpublished paper at the recent IFLA Congress in August 2004 "E-Thesis and electronic publishing: a strategic position for university libraries" by Jean-Claude Guedon (Université de Montreal, Canada) reported that digital repositories had not taken off. His feeling was this was mainly due to academics being nervous about the regulations allowing them to place their papers on a repository and secondly feeling that there was no point in putting papers up that had not been refereed. Therefore he proposed that in order to get experience, librarians should trial the repositories on theses and dissertations. I did not get chance to ask a question but this would have been about the problems that librarians might have in doing theses rather than the other kinds of materials. Librarians do not generally control the presentation of theses and dissertations in universities. They would need to negotiate with the departments that do (university registries or academic departments) before making electronic theses and dissertations available in this way. Anyway the gist of the paper was clear, that it has not been easy to set up digital repositories.

The other key event in the UK has been the publication of a report by the Science and Technology Committee of the British House of Commons entitled *Scientific publications: Free for all?[v]* This has been occasioned by the progressive increase in prices of journals particularly those that are available in digital form. There is a feeling that journal publishers are taking advantage of the monopoly situation to charge what they wish. Why is it a monopoly situation? If you look at the price of canned fruit you will see a wide range of products and varying prices. If you want quality or low price you can take your choice. You can even have oranges instead of peaches. In the case of journal articles, the situation is completely different. A scholar or researcher will need a particular article if it exists. There is only one way to get that article usually. Indeed the article should not appear more than once since it can only be published in one journal. The publishers can therefore charge as much as they think the libraries can afford. even more, since they could argue that their digitisation could enable libraries to employ fewer staff and thereby afford to pay more for their journals

The report came down very much on the side of libraries. For example it stated that academics are unwilling to ignore the problem faced by libraries. This is the problem that periodicals are becoming more expensive, there are more of them, universities do not have any more money and academics do not want any of their funds to go to the libraries to pay for the publications they want the libraries to obtain. They merely complain that the library does not buy enough journals. The report includes a substantial section on institutional repositories, and explains the role that Professor Stevan Harnad has had in fostering them as part of the Open Access movement (the other being Open Access journals). It speaks of the SHERPA project and the FAIR programme. It also talks of the possibilities for the use of repositories, and the impediments. One is that if an article is published elsewhere it cannot be published in a journal. Another is that the idea of a repository cannot be realised if journal articles cannot be placed in it. Imagine this scenario. A university lecturer writes an article publishes it in a journal which the university library cannot afford. The lecturer wants his students to read an article he has written. The copyright of this belongs to the journal publisher. If he wants his students to read this they must buy the journal or go to another library that cam afford it. The teacher can put the article in a repository but he must seek permission from the publisher who may not grant it.

The intellectual property is the lecturer's but it is taken at no cost by the journal publisher. This must be wrong. Publishers have seen the bad press they will get and one large one, Elsevier, changed its policy just before the report was published. It now allows the text of articles to be reproduced as text only and on private, though not central, servers. However, this is not satisfactory as it does not allow diagrams, illustrations etc to be reproduced. Open access journals are another alternative. Here the costs of the journal are paid for by the author paying to publish. But the journals are free to read. In effect this means that a research project pays. Many academics feel that this merely transfers the cost of journals from the library to the teaching and research budget.

The report heard that the University of Hertfordshire (the neighbour of Middlesex University and of a similar level) was waiting to see if other universities set up a repository before setting one up itself. The report suggested that there was no need to wait and recommended that:
 a) all publishers allow their authors to archive their articles in a repository
 b) all UK research institutions establish and maintain repositories and
 c) all UK academics deposit articles in their institutional repositories

For self-archiving, the Committee specifically recommended that:

a) universities should be funded to provide institutional OA e-print archives,

b) authors should self-archive their papers within a month of publication,

c) funding councils should mandate that all their funded work is self-archived and

d) the British government should act as an agent for change both in the UK and in the international arena.

5.The future

What could be better for the open access concept. All we are waiting for is for government to give money to libraries to achieve this! Nevertheless the report recognised that academics are reluctant to use repositories. Most universities like to see their academics publishing in top level journals. There is a conflict here between the traditional methods of securing academic recognition and the need to take advantage of modern technological possibilities for disseminating knowledge. Government support brought on by the need to escape from the ever-increasing spiralling of costs of publishing may well change that and there may not be seen to be this kind of conflict in the future. Most probably the two systems will go side by side. Methods will be required to ensure the quality of material that is archived. In fact this can be done internally since university departments will not want to gain a reputation for archiving poor quality papers. They will want to set up boards probably with external members to assure the level of quality is maintained. But we are at the start of a new era and no one can be sure which way events will turn.

References

----

[i] http://www.eprints.org

[ii] Ginsparg, Paul. First steps towards electronic research communication. Computers in physics 8(4) Jul. 1994, pp390-396

[iii] http://www.soros.org/openaccess/read.shtml

[iv] http://www.sherpa.ac.uk/

[v] House of Commons Science and Technology Committee. Scientific publications: Free for all? Tenth Report of Session 2003-04. London: HMSO, 2004