

Establishing a National Network of Repositories

This paper addresses a number of specific questions forwarded by the Science and Technology Select Committee to the authors on 28 April 2004.

How much, approximately, would it cost to establish a national network of repositories?

The easiest way to establish a national network of repositories would be for each higher education institution (HEI) to set up its own repository or set of repositories. The cost of setting up an individual repository for each institution is small. The software is free – there are several packages available, all of which are freely downloadable¹. The software can be installed on a standard server, costing about £1,500. It takes a computer officer between 2 and 5 days to get it all working (at a cost of say £600 for the time).

Experience from the SHERPA project shows that further customisation work is required during the first 6 months of the repository's life. This includes developing the 'look and feel' of the service, and ensuring it is fitted into the support and maintenance procedures of the institution. We estimate this adds up to 15 person days' work over the course of 6 months. A cost of about £1,800. Thereafter, ongoing technical work can normally be absorbed within an institution's standard maintenance routines.

Populating the repository has its own costs. At present, the major costs are in advocacy – promoting the service and persuading academics to deposit articles in the repository. This cost would, of course, be removed if there was a mandate (from institutions, research councils, or other agencies) that publicly-funded research had to be made publicly available in this way. Research shows that most academic authors would be happy to deposit their materials in a repository if this kind of mandate was introduced. The costs for depositing items are given below.

The overall cost of setting up a national network of institutional repositories would therefore be the costs for each institution multiplied by 131 (the number of HEIs in the UK). See below for a summary of costs. It should be noted that some institutions already have repositories or have plans to install them.

How many repositories would the UK need in order to achieve maximum coverage?

Since the entry barriers (including costs) are very low, the normal assumption is that every institution which produces research publications should have a repository. Most institutions would prefer to have their own repository which they can set up in line with local requirements. The repository can act as a research management tool and as a 'shop window' for the institution's output. Most institutions already have the organisational and technical

¹ For example, EPrints.org software - see <http://www.eprints.org/>
Establishing a National Network of Repositories -
Supplementary Evidence for the Parliamentary Inquiry into Scientific Communication

infrastructures to set up and maintain a repository easily. However, some institutions might for various reasons choose to join a consortium with a shared repository. Both models (individual institutional repositories and consortial repositories) are being tested in the SHERPA project². In addition, the British Library, as a SHERPA partner, is setting up a repository which will hold content from non-affiliated researchers.

A network of institutional repositories would, of course, operate as a single virtual archive. When repositories are set up using the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting³), they can be searched in a seamless way. Users may not even be aware that they are searching the contents of multiple repositories. The actual location of the content is in this sense irrelevant, as long as it is in an OAI-compliant repository. Once the material is available online, then other services can use it – to present subject-based collections, for example, or as a source material for teaching. Repositories run by institutions will provide a stable foundation on top of which other services can be built. Research results can be consulted and also re-used in new ways to enhance the return on the original research investment.

Having a repository in each research-active HEI plus the British Library repository would give coverage for most UK research output.

What level of resourcing would be required to maintain the repositories?

The technical maintenance costs of a repository would be minimal. Most could be easily absorbed into an HEI's standard web services maintenance. Every 3 years or so there would be a need to replace the kit and possibly upgrade the software but the costs of this (hardware plus 2 to 5 person days) would be no more than initial installation.

Populating the repository would be an ongoing cost. Two scenarios are possible here. The first is that authors archive their own material as and when it is produced. This is a simple process which takes at most about 15 minutes for each paper using a specially designed web interface (available as standard in repository software). In this case, the costs are effectively hidden as they are distributed amongst a large number of researchers who each spend a small amount of time (1 to 2 hours per year) depositing their own papers.

The second scenario is that someone is employed to deposit items on behalf of researchers. Most research-led HEIs would require a full-time member of staff, able to work across subject-disciplines and co-ordinate the collection of material. The cost of such a post would be up to £30,000 per year. An alternative might be that departmental administrative support staff could carry out the role for their department. In this case, the amount of work would be distributed and most of the costs would again be absorbed by the institution.

² JISC funded project to co-ordinate the building of institutional repositories in project partners' institutions - see <http://www.sherpa.ac.uk>

³ Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) defines common standards for storing and searching material in repositories - see <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Establishing a National Network of Repositories -
Supplementary Evidence for the Parliamentary Inquiry into Scientific Communication

In reality, a mixed economy may develop, with a variety of academic authors, research assistants, departmental officers and library personnel depositing papers.

In the longer term (20 to 50 years), the big costs are likely to be the preservation of the digital files. These costs are, of course, not unique to repositories – they apply to all electronic resources intended for long-term use (including those currently produced by commercial publishers). Although some work has been done on digital preservation costs (by the British Library amongst others), this is still a rather unknown quantity. It is a problem that needs addressing for all digital objects held in repositories and everywhere else. It need not stop repositories being set up and populated in the short term, however, as the issue – and costs – will need to be faced at some point whether the material is held in institutional repositories or elsewhere.

How practicable would it be to establish a network of UK repositories?

Very practicable. Establishing the network would be low cost and the expertise is already there in most institutions to do it without delay.

It is already happening. In a recent informal survey of members of SCONUL (the Society of College, National and University Libraries) we found that of the 51 respondents, 17 (a third) already had set up repositories (mostly research-led universities). A further 13 were expecting to do so in the foreseeable future. This means that about 59% of the respondents already had a repository or expected to have one in the near future.

The main challenge at the moment is not setting up the repositories *per se* but populating them. Academics do not currently have many major incentives to archive their material (or at least they are unaware of the benefits of repositories). Mandating the depositing of research papers produced by HEIs in open-access facilities would be a practical way to overcome this problem. It would mean that academics or their institutions would be required to retain sufficient rights in their work to deposit it in repositories.

Summary of costs

Costs are given per institution (there are 131 HEIs currently in the UK). *It is assumed that many of the maintenance costs of institutional repositories can be absorbed by institutions.* The costs given are minimal; higher costs would, of course, apply if institutions wished to enhance the service in various ways.

Installation costs:

Server	£1,500
Software	£0
Installation (5 days)	£600
Customisation (15 days)	£1,800
Total per institution	£3,900

Ongoing maintenance costs:

Technical support	Absorbed by institutional IT services
Supported archiving service	£30,000 per year
Upgrades/migrations	£3,900 every 3 years
Digital preservation objects)	Significant costs (applies to all digital

Stephen Pinfield
Bill Hubbard
University of Nottingham
12 May 2004