



A NATIONAL MODEL FOR SHOWCASING RESEARCH

Most countries have a number of Open Access repositories (archives) maintained by individual institutions or organisations. Repositories may include those at universities and research institutes, there may be cross-institutional repositories and repositories may also be maintained by publishers. This is the sort of pattern seen in many nations now and such a network of repositories forms the basis for a 'shop window' for a country's research.

All these repositories are indexed by Google, Google Scholar and other web search engines, so their contents can already easily be discovered by using these services. Sometimes, however, countries wish to provide a service that actively showcases their national research output to the rest of the world, rather than just leaving people to find the nation's research via web search engines.

What are the technical issues involved?

It is important at this point to note that a research paper deposited in an electronic repository (archive) is composed of two main elements: first, the **metadata** and, second, the **full-text** of the article itself and that these can be treated as separate entities (see sidebar for further explanation).

The other important thing to note is that repositories operate according to international technical standards, meaning that their contents share a basic structure and can be harvested by service providers wishing to include that content in their service. So, for instance, the *Economists Online* service, which showcases the economics research literature, harvests the metadata of tens of thousands of economics articles from repositories and brings them together in one big collection (1). The service adds lots of useful features for users and makes a great showcase for the economics literature, yet the actual full-text articles remain in lots of different repositories distributed around the world. Only the metadata are collected together in one place to make the service: when a user finds an article of interest there is a link from the metadata to the full-text in its original location in a university, research institute or similar repository.

What should a national system look like?

The question is how best to organise a national-level system. The main issue is where authors should be required to deposit their work. A national service may require authors to deposit their articles directly in a central database. This is an inadvisable strategy because it imposes a double burden on authors whose institution already has a repository in which they are required to deposit their papers. They do not want to go through the deposit process twice or more.

A study was commissioned by the UK's Joint Information Systems Committee to determine ways to provide a national view on research and its findings are helpful for other countries, too (2, 3). The study describes the three main ways to provide a national-level service. These are:

Centralised model: under this model both the metadata and the full-text articles themselves are submitted *directly* to a central database by the authors.

Distributed model: under this model all metadata and articles are collected in institutional archives (e.g. university repositories) and these are searched 'on the fly' when a user looks for an article

Harvesting model: this is a hybrid model. Under this model, the articles are collected in institutional archives and the metadata are harvested into a central searchable database. It is also possible to extend this model to one where the full-texts are harvested too and enhanced in various ways, such as using semantic technologies on them to provide a better quality search facility or to provide additional information about, and from, the content. The critical point is that the articles are only ever **deposited** once, in institutional repositories, and thereafter any services that wish to use them harvest what they need from these repositories.

An electronic research article consists of two main elements:

* **Metadata:** these are the things that **describe** the article. They include the names of the author(s), the author affiliation, the title of the article, the journal name and issue number (if known) and so forth

* **Full-text:** this is the text of the article along with its images, graphs, diagrams and tables: in other words, the full document that people may want to read

In addition, there may sometimes be **supplementary material**, such as datasets or other accompanying files. The metadata can be copied and stored in a separate place from the full-text of the article itself and yet remain electronically linked to it.

Why is the harvesting model the best one?

This harvesting model was found to be the optimal model for a national-level approach to providing a service that presents and showcases research, either research in a particular field or even the whole output of a country. The advantages of this model over the others are that:

- The articles have been deposited once already in an institutional repository and the service simply harvests their metadata into the national collection and points would-be readers back at the full-text in the institutional repository. There is no need for second or multiple deposits to be made. Authors like this solution
- Institutions like the solution too, since their interest is in collecting together the research outputs from the institution and they do not want their authors discouraged by having to undertake the deposit step more than once. Institutions can also monitor and manage the process to make sure that all their research is being deposited, which a centralised deposit facility cannot do
- It is a low-cost option to operate since storage requirements are minimal and the metadata-harvesting technology is simple
- It provides much scope for the service provider (the central database) to add features and enhance the offering to users if funds are available
- It works equally well, technically, for articles in repositories, in open access journals, or elsewhere
- It capitalises on the already-established and growing number of institutional repositories that are collecting research together at institutional level. The model simply aggregates those local collections into one big national one

Working examples of national services

The UK's *Intute Repository Search* service (3) provides such a viewpoint for the UK's research outputs - a shop window for British research. *Intute* harvests the metadata, and in some cases the full-text, of articles in British repositories and shows it off through a smart, easy-to-use service that has many user-friendly features that Google, and other straightforward web search engines do not provide. The other new UK service is **EThOS (Electronic Theses Online)** (4), a service run by the British Library showcasing British masters' and doctoral theses.

Showcasing research

With growing numbers of institutional repositories, and growing amounts of content in those repositories, the opportunity is now here for any country to showcase its research effort. National-level service(s) can cover all research, as *Intute* does in the UK, or provide services by subject area, as *Economists Online* does. Or a service might harvest and showcase masters' and doctoral theses, like the UK's EThOS service does.

Existing institutional repositories - and many more are expected to be established over the next few years - contain between them millions of articles. Using these to establish national-level services will help to optimise the visibility of any country's research and increase its impact.

References and further information

1. Economists Online: http://www.neeoproject.eu/about_neeo.html
2. Swan, A., Needham, P., Proberts, S., Muir, A., Oppenheim, C., O'Brien, A., Hardy, R. and Rowland, F. (2005) Delivery, Management and Access Model for E-prints and Open Access Journals within Further and Higher Education. Technical Report for the JISC. <http://eprints.ecs.soton.ac.uk/11001/>
3. Swan, A., Needham, P., Proberts, S., Muir, A., Oppenheim, C., O'Brien, A., Hardy, R., Rowland, F. and Brown, S. (2005) Developing a model for e-prints and open access journal content in UK further and higher education. *Learned Publishing*, **18** (1), pp. 25-40. <http://eprints.ecs.soton.ac.uk/11000/>
4. Intute Repository Search <http://www.intute.ac.uk/irs/>
5. EThOS (Electronic Theses Online) <http://www.ethos.ac.uk/>



This Briefing Paper is one of a series provided by **Enabling Open Scholarship (EOS)** the organisation of universities and research institutes promoting open scholarship around the world.

EOS provides information, help, guidance and a forum on open scholarship issues for senior managers. For further information, see www.openscholarship.org