



What are institutional repositories?

Institutional repositories are digital collections of the outputs created within a university or research institution. Whilst the purposes of repositories may vary (for example, some universities have teaching/ learning repositories for educational materials), in most cases they are established to provide Open Access to the institution's research output and this is the focus here. A short account of the development of institutional repositories can be found at <http://poynder.blogspot.com/2006/03/institutional-repositories-and-little.html>.

How many are there and where are they?

There are over 1300 repositories around the world at the beginning of 2009. Over the past three years the number has been growing at an average rate of one per day. The statistics on numbers and where they are can be found in the Registry of Open Access Repositories (ROAR: <http://roar.eprints.org/>) and in the Directory of Open Access Repositories (OpenDOAR <http://www.opendoar.org/>). There is also a map showing where they are at Repository66 (<http://maps.repository66.org/>).

The advantages of a repository to an institution

A repository has the following purposes and benefits for an institution:

- * Opens up the outputs of the university to the world
- * Maximises the visibility and impact of these outputs as a result
- * Showcases the university to interested constituencies – prospective staff, prospective students and other stakeholders
- * Collects and curates digital outputs
- * Manages and measures research and teaching activities
- * Provides a workspace for work-in-progress, and for collaborative or large-scale projects
- * Enables and encourages interdisciplinary approaches to research
- * Facilitates the development and sharing of digital teaching materials and aids
- * Supports student endeavours, providing access to theses and dissertations and a location for the development of e-portfolios

How repositories help Open Access

Repositories adhere to an internationally-agreed set of technical standards that means that they expose the metadata (the bibliographic details such as author names, institutional affiliation, date, titles of the article, abstract and so forth) of each item in their contents on the Web in the same basic way. In other words, they are 'interoperable'. This common protocol to which they all adhere is called the open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The contents of all repositories are then indexed by Web search engines such as Google and Google Scholar, creating online Open Access databases of freely-available global research. As the level of self-archiving (the process by which authors deposit their work in repositories) grows the Open Access corpus will represent an increasingly large proportion of the scholarly literature.

What do institutional repositories contain?

Peer-reviewed journal articles and conference proceedings

The primary type of content in repositories is the peer-reviewed journal literature. A collection of the journal articles published from an institution, provided in Open Access through the repository, gives the institution's research programme worldwide visibility and increases its impact. Individual authors also enjoy the same increased visibility for their work and concomitant impact.

Research data

Now that research data are increasingly created in digital form, repositories are also places where authors can deposit the data that underpin their final articles. More and more research funders are requiring their grant-holders to make their data Open Access, once they have themselves analysed and published their findings from the data. This is in order that other researchers can use the data to verify results, to compare

with their own data or to re-use in some way to generate new data and knowledge. Datasets may be of many types – spreadsheets, photographs, audio files, video files, representations of artwork, diagrams, charts and so on. They may even be ‘complex objects’, that is, combinations of several types of data, such as a numerical dataset recording weather patterns with accompanying satellite images.

<http://www.dlib.org/dlib/september07/treloar/09treloar.html>

Monographs and books

Most institutional repositories also contain books or book chapters. Books are often written for monetary gain (royalties on sales) and in such cases authors may be reluctant to deposit them for free in a repository. In these cases it is still important for the book to be deposited, with the metadata (title, author, synopsis, publisher details, etc) on display, but the text may be ‘hidden’ from viewers. Having the metadata visible means that the book is counted in the institution’s assessment procedures, its existence is known to would-be readers and it can be located by Web search engines. The evidence is accumulating, however, to show that when the entire content of a book is visible in a repository, sales of the book frequently rise. This is because the visibility in the repository is raising awareness of the book and promoting it to an audience which is then likely to buy the book if it seems relevant to their work. It is analogous to what Amazon offers with its ‘Look inside’ facility.

Other content types

As well as the types of content described above, institutional repositories frequently contain theses, dissertations and other research-related outputs such as presentations.

How do repositories fit into the scholarly communication landscape?

Repositories will form a permanent and critically important part of the scholarly communication process. Their first role is to provide the Open Access literature. Additionally, services may be added to repositories to provide extra functionality. For example, a usage-reporting service gives authors and the institution information on how the content of the repository is being used. A search service may help users find specific items more easily. A service that organises content in specific ways may help authors, for example, to download a list of articles into their CV, or aid institutions in assessing the institution’s research programme or for reporting data to governments or for other statutory requirements. We may be looking forward to a time when repositories play a formal role in the publishing process. Repositories can collect articles from the institution’s authors when they are ready for peer review and a peer review service will collect them from the repository for processing. There are already signs of these things happening. A few scholarly society publishers encourage authors to notify them when a paper has been deposited in a repository and is ready to be peer reviewed and published. Some university presses are working hand-in-hand with the repository when publishing books by institutional authors.

Who uses institutional repositories?

Because Google and the other Web search engines index the content of repositories, anyone with internet access can find themselves arriving at an article or dataset in a university or research institution’s repository via a Web search. But there are other ways that repositories are used, too. Users may search a particular repository if they know a specific researcher works at that institution. Or they may follow a link from another researcher’s website or blog. Although these specific ‘referrals’ are not uncommon, by far the most common way for researchers to arrive in a repository is through a Web search engine such as Google. Les Carr’s data (<http://listserver.sigmaxi.org/sc/wa.exe?A2=ind06&L=american-scientist-open-access-forum&D=1&F=PI&P=16792>) on how the repository at Southampton University is used showed that Web search engines accounted for 64% of user traffic into the repository. This underlines how important these informal ‘world research databases’ that the Web search engines have created are for repositories and their institutions.

Sustainability of repositories

Repositories represent a certain basic cost to an institution but there are ways of minimising this. The sustainability of a repository depends in large amount on how much effort is needed to fill it. A repository that is filled by self-archiving – that is, by researchers depositing their own articles – is far less costly to an institution than one where the library does all the depositing work. Ideally, deposit activity should show a reasonably steady pattern throughout the year (<http://eprints.ecs.soton.ac.uk/13872/>).

The way to ensure that researchers deposit their work regularly is to have a proper Open Access policy which requires them to make their work Open Access and explains why this is important for them and the institution.