

# E-Prints Digital Repositories Software

## Digital Repositories and Digital Curation

Digital Repositories play a vital role in the curation of digital materials and offer a convenient way to store, manage, reuse and curate a variety of digital materials. The term digital repository can be applied to a number of different digital storage initiatives, which are often also referred to as 'institutional repositories', 'digital archives', or 'digital libraries' although in practice these each have slightly different functionality and underlying philosophies.

A growing number of repository models and systems are available and used by a variety of communities. They can take many forms and carry out many different functions. This technology watch paper provides an introduction to the features and functionality of the EPrints digital repository system.

## EPrints

The EPrints software package was originally designed for creating and managing open access institutional repositories of research papers and publications. It is widely implemented and is now used to store and manage a much broader range of content types and for different purposes.

EPrints software is managed and developed by the Electronics and Computer Science department at the University of Southampton and is freely available as open source software under the GNU General Public License (GPL). The GPL allows users to freely run, study, modify, improve, and release improvements of the program under the same terms. Most of the installation process is automated by way of an installation script and further help is available from the website. The default configuration is oriented towards research papers; using the system to manage different types of digital materials – such as audio archives and research data – requires changes to the configuration. EPrints is written in Perl, requires installation of the Apache web server and a number of Perl modules, and uses the MySQL database as a back-end. As with EPrints, all of this additional software is open source and freely available. It was developed under GNU/Linux and is intended to run on any UNIX-like system. There are no plans for a version to run under Microsoft Windows. Online support and advice is available from the EPrints wiki and mailing lists, although a charge is levied should implementers choose to join the EPrints community or solicit personal support services from the EPrints team.

EPrints is freely distributable and, in keeping with the terms of the GPL, the source code is open and freely modifiable by programmers on the condition that modifications are also free and open.

## Selected Implementations

EPrints has been installed and is running in over two hundred repository systems worldwide. Most repositories hold only electronic publications, but a number have been developed specifically to support research output. These include the Southampton Crystal Structure Report Archive (developed as part of the UK eBank project), the National Aerospace Laboratories Institutional Repository, and the IUBio Software Archive (a heavily modified version of EPrints).

The University of Tasmania (AUS) has developed several open source EPrints extensions that enhance the base functionality of the system. One such extension allows depositors to see how many times their work has been accessed, which may be useful in encouraging authors to deposit. The DAEDALUS project developed Perl scripts for the University of Glasgow EPrints Service to import bibliographic details from other applications and databases. These are also freely available.

# Functionality

As development of EPrints is ongoing, only initial and base functionality is discussed here. This paper refers to EPrints V2.3.

**Ingest.** The EPrints development team are proponents of open access and 'self-archiving', a deposit philosophy whereby authors deposit their own works and metadata in a preferably OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) compliant EPrints repository. The deposit process reflects this aim, as it is very straightforward to encourage authors to deposit their works (although in practice, intermediaries are often the main depositors). Depositors must register to upload document files and metadata to the EPrints repository via the web-interface, although some institutions offer a 'proxy self-archiving service' and allow depositors to e-mail relevant documentation to repository staff who upload the item and metadata on the depositors' behalf. EPrints has a bulk data import facility that can be used to import XML files.

The default deposit configuration requires depositors to enter a minimum set of metadata. This set does not conform to any particular standard, although in practice, an EPrints implementation can be configured to request and store any metadata according to institutional requirements and the type of objects stored. A Dublin Core application profile for deposited ePrint items is under development as part of the JISC Digital Repositories Programme.

**New objects** (i.e. content file(s) + metadata) are temporarily stored in the registered user's 'workspace' until all required metadata has been entered and the object deposited: this enables depositors to put an entry 'on-hold' until all required metadata has been collected. A built-in buffer enables items to be withheld for review or approval by repository administrators before final entry into the system. This can be disabled.

**Workflow.** In keeping with the straightforward ingest procedures, workflow functionality is minimal. Editors and administrators can receive e-mail notification when new items are deposited and move content in and out of the submission buffer if the buffer is active.

**Storage.** MySQL must be installed as the back-end Relational Database Management System (RDMS) for an EPrints service. The amount of discspace required depends on the repository scale, but approximately 2MB should be allowed for each ePrint item deposited. Content can be stored in any format designed acceptable by the repository administrator during configuration. Multiple representations of the same content are acceptable.

**Preservation.** Functionality for long-term preservation is not currently an explicit feature of the EPrints architecture. Preservation in institutional repository software, particularly in open access, is a source of some contention, with conflicting opinions on whether it is the role of an institutional repository to undertake long-term preservation of journal and research articles that have been published elsewhere. Despite this, the EPrints team is participating in the PRESERV project to investigate and develop infrastructural digital preservation services for digital repositories.

**Access.** A web accessible interface enables users to search the system. Full text searching and searching on metadata fields are both possible. The software does not assume open access and can be configured to allow limited access to certain deposits or sets of deposits. The generic version of EPrints is fully interoperable with all other OAI-PMH compliant Open Archives, which means that papers in all OAI-PMH compliant Archives can be harvested using the OAI-PMH protocol.

---

## Additional Resources

EPrints website: <http://www.eprints.org/>

Hunter, Philip, *ePrints UK Final Report (2005)*: <http://www.rdn.ac.uk/projects/eprints-uk/docs/final-report/eprints-uk-final-20050316.pdf>

Suber, Peter, *Open Access Overview (2006)*: <http://www.earlham.edu/~peters/fos/overview.htm>

DAEDALUS project bibliographic scripts: <https://dspace.gla.ac.uk/handle/1905/175>

Dublin Core EPrints Application Profile development:

[http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Application\\_Profile](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile)

PRESERV (Preservation EPrint Services project): <http://preserv.eprints.org/>

University of Tasmania EPrints extensions:

<http://leven.comp.utas.edu.au/AuseAccess/pmwiki.php?n=Software.EPrints>

UK eBank project: <http://www.ukoln.ac.uk/projects/ebank-uk/>