Eprints Institutional Repository Software: A Review

Mike Beazley
Academic librarian
Vaughan Memorial Library, Acadia University
mike.beazley@acadiau.ca

Keywords
Institutional Repository, Eprints, Review, Software, Open-Source

Abstract
Setting up an institutional repository (IR) can be a daunting task. There are many software packages out there, some commercial, some open source, all of which offer different features and functionality. This article will provide some thoughts about one of these software packages: Eprints. Eprints is open-source, and the software is easy to modify. This presents clear advantages for institutions with smaller budgets and that have programmers on staff.

Installation and initial configuration are straightforward and once the IR is up and running, users can easily upload documents by filling out a simple web form. Eprints is an excellent choice for any institution looking to get an IR up and running quickly and easily, although it is less clear that an institution with an existing IR based on another software package should migrate to Eprints.

Introduction

Setting up an institutional repository (IR) can be a daunting task. There are many software packages out there, some commercial, some open source, all of which offer different features and functionality. This article will provide some thoughts about one of these software packages: Eprints. Eprints was one of the first IR software packages to appear and has been available for 10 years. It is under continual development by its creators at the University of Southampton and the current version is v3.2.3.

Eprints is open-source, meaning that anyone can download and make use of the software for free and the software can be modified however the user likes. This presents clear advantages for institutions with smaller budgets and also for institutions that have programmers on staff. Eprints requires some additional software to run: Linux, Apache, MySQL, and Perl. This software is all open-source and already present on the servers of many institutions. There is now a version of Eprints that will run on Windows servers as well, which will make the adoption of Eprints even easier for some.

In brief, Eprints is an excellent choice for any institution looking to get an IR up and running quickly and easily. Installation is straightforward, as is the initial configuration. Once the IR is up and running, users may upload documents and provide the necessary metadata for the records by filling out a simple web form. Embargoes on published documents are handled
elegantly by the software, and the software links to the SHERPA/RoMEO database so authors can easily verify their rights regarding IR submissions. Eprints has some drawbacks, which will be discussed later in the review, but on the whole it is easy to recommend to anyone looking to start an IR. It is, however, less clear that an institution with an existing IR based on another software package should migrate to Eprints.

Features of Eprints
The following is a list of some of the features of Eprints that illustrate its potential as an IR software package. A more complete list of features can be found on the Eprints website: http://www.eprints.org/

- Easy end-user uploads.
- Bulk importing and exporting of records (ASCII, BibTex and more) for uploading established collections.
- Three user roles: administrator, editor and author.
  - Administrator role controls all back-end options such as organization of records, web interface appearance and functionality, and all other server-side settings.
  - Editor role reviews submissions before they are published online and may edit metadata on submissions to maintain consistency or correct errors.
  - Author role allows submission of documents and management of previously submitted documents.
- Easy search and browse features (multifaceted browsing available, customizable by administrator).
- All necessary software for full functionality is open-source (Linux, Apache, MySQL, Perl).
- Provides RSS feeds for entire collections or based on specific criteria such as subject, author, etc.
- Functions with many file types, including: PDF, HTML, JPEG, TIFF, MP3, and AVI.
- Thumbnail preview of documents and images is generated automatically upon file upload.
- Eprints has a 10 year history with an active development community.
- Easy to develop plug-ins using Perl.
- OAI compatible (which means that Google Scholar can index the contents of your Eprints archive).
- Any materials that cannot be displayed online can be requested with the click of a button.
- Temporary restrictions to accommodate embargo periods are easy to set.
• Preset and custom authority files available to maintain metadata consistency and avoid ambiguity (with author names, for example).

• Custom subject categories for browsing (faculty, department, LC subject headings, etc.).

• Integration with SHERPA/RoMEO for quickly checking publisher policies and author rights.

Existing Eprints Community

Eprints is currently in use in at least 269 institutions worldwide. These institutions include: Australia National University, Birmingham University, CalTech, Harvard College, and the British Library. Many of the Eprints implementations currently contain several thousand records, with a few implementations containing over 10,000 records (CalTech is an example of such an implementation). The largest known Eprints repository is housed at the University of Twente in the Netherlands. This repository contains over 60,000 records, demonstrating the ability of the Eprints software to handle very large collections ("Open Access"). As Richard Wyles points out, Eprints has an install base that is larger than many of its competitors, which suggests that the Eprints software is indeed easy to implement in a wide variety of institutions (9).

Strengths of Eprints

The real strength of Eprints lies in its ease of use for both end-users and administrators. Submitting documents in Eprints is very straightforward. Users are taken through the submission process one step at a time and asked to provide metadata information along with an electronic copy of the document. Users simply enter metadata such as document type, title, author name, date, etc. via a web form, so no knowledge of HTML or XML is required. The metadata fields that appear in the form are selected by the administrator. This form is easily customized by the administrator, so that only the fields that are pertinent to a given collection are presented to the end-user. Users can manage their submissions to the archive, and editing, updating, and removal of documents is possible after submission (although the administrator can limit these functions).

Another nice submission feature is the ability to specify an embargo period before a document is made available in the archive. If an embargo period is necessary, the user may submit the document at any time and simply specify on what date the embargo ends. When that date arrives, the document will automatically be made available in the archive; no further input from the user or the administrator is required.

Although Eprints does not support boolean search, it does provide multiple search options as well as a robust and highly customizable browse feature for finding documents in the archive ("Repositories Support Project"). Browsing can be done based on any of the metadata fields within a collection, and multiple browsing criteria can be used. For example, in browsing a collection of theses, it would be possible to browse by department and then break down the results by supervisor and year. The administrator has control over which...
browsing categories are made available to the user. Eprints is OAI-compliant, and documents in an Eprints archive can be indexed by Google, which helps to ensure greater access to, and greater dissemination of, any items uploaded to the archive.

The initial installation and configuration of the Eprints software is advertised as being quick and uncomplicated ("Eprints"). This assertion is backed up by user feedback and other published reviews (Wyles 9; Stranger and McGregor 3). For institutions without the necessary technical expertise, Eprints Services, a company created by the developers of Eprints, can help with installations and configuration and can even arrange hosting of an Eprints archive off-site.

**Weakness of Eprints**

While the benefits of Eprints outweigh its shortcomings, there are a few drawbacks that should be noted. The biggest of these drawbacks is the absence of a bulk uploading feature. Uploading files and creating records is certainly simple in Eprints, but if one needs to upload a pre-existing collection, there are no options available for creating multiple records at once. Multiple files can be uploaded at once, but only when attached to a single record. It should be noted that if your institution wishes to transfer records currently held in other IR software, Eprints can easily import these records, so the inability to perform bulk record creation is only a problem for collections that are not currently contained within a database. However, if your institution wishes to make, for example, a photo collection available online, each photo must be uploaded and each record created individually. To provide a concrete instance when this may pose a problem, one might think of a collection of photos of a particular type of animal—say, horses. These photos would require certain common metadata tags, most obviously the general subject heading, which in this case would be horses. In Eprints, the user would not be able to upload the whole photo collection, applying one or more pieces of metadata to all photos and then add unique metadata to individual photos within that collection afterwards. This would mean that if the collection of horse photos contained various breeds of horses and were taken at different locations on different dates, this data could not be individually applied to pertinent photos. Ultimately, in Eprints your options would be either to have a generic record for multiple photos or to devote the time to creating individual records for each photo one at a time.

The other drawback to Eprints is its search function. The search features in Eprints are fairly limited. As mentioned previously, boolean searching is not supported. It is also quite easy to run a search that yields zero results. For many end-users accustomed to modern search engines and databases, an unsuccessful search (with no suggestions for alternative search strategies) may be discouraging. In addition, Eprints does not allow user-creating tagging, a feature which some modern databases, including newer library catalogues, often provide. While the robust and customizable browse feature mitigates some of these problems, a more sophisticated search feature would be welcome in future iterations of the software.
Conclusion and Recommendation

Eprints seems to be a very good system for hosting an incrementally growing collection, such as a faculty publications collection or an electronic theses collection. The ease with which Eprints allows for document submission is a boon to any institution wishing to instil a culture of self-archiving among its members. It also removes the need for staff or administrators to handle the file uploading and data entry that accompanies every document submission. Control of what is made available online can still remain in the hands of the administrators, because all submissions must be approved by an editor. This allows for the institution to maintain some uniformity regarding metadata as well as to ensure that submissions fall within the scope of the archive.

Eprints would be a highly suitable solution for developing various collections that depend on submissions from disparate sources. The thrust of the marketing for Eprints strongly emphasizes user-submitted content rather than the management of large collections already held by an institution. Management of existing collections is a well advertised strength of many competing IR packages and thus, abandoning any currently used IR software in favour of Eprints may not be worthwhile when dealing with large, established collections. However, Eprints is an excellent option for the quick and easy development of user-driven collections within any institution.
Works Cited


