

Eprints.org Software: a Review

Ed Sponsler and Eric F. Van de Velde
California Institute of Technology
July 27, 2001

Stevan Harnad just made it easy to join him in his quest to free the scholarly literature. Download Eprints software from <http://www.eprints.org/>, and build your own repositories quickly and easily. Developed by Harnad collaborators Robert Tansley and Chris Gutteridge, Eprints provides a web interface for managing, submitting, discovering, and downloading documents. Eprints repositories are compliant with the Open Archives Initiative (OAI) (<http://www.openarchives.org>). Therefore, once a repository is registered as an OAI data provider, OAI-aware information services will be able to discover its content.

This review is based on our eight-month experience with Eprints. During this time, we built several technical-report repositories and one online conference proceedings. Our repositories are available at <http://library.caltech.edu/digital>.

Setup, Configuration, and Administration

The cost of getting started is minimal. An experimental system, suitable for initial testing and even for hosting a few small production repositories, costs less than \$1000. Obviously, one must move up to higher price and performance points when the number and size of the repositories, the number of users, and the performance requirements increase.

Eprints requires the Linux operating system with a standard configuration of supporting software:

- Apache, the web server
- MySQL, the relational database
- Perl, the scripting language
- Various plug-in modules for Perl

The operating system and all of the supporting software are Open Source software licensed under the GNU General Public License (GPL). (See <http://www.fsf.org/copyleft/gpl.html>.) Eprints developers intend to make Eprints officially Open Source as soon as they have implemented certain critical features. (Private communication) In the mean time, the University of Southampton holds the copyright, but it allows users to view, modify, and redistribute the source code. This is very close to being Open Source in the GPL sense. (See <http://www.eprints.org/download.php> for full details.)

After installing the basic software, one must configure the Eprints system for local use:

- Customize the look and feel of the local Eprints web site by adapting scripts that control the presentation. These scripts are well separated from the core Eprints code that deals with archiving, database management, and internal workflow. Therefore, we expect that future upgrades will leave the customized scripts largely unaffected.
- Decide what metadata fields to use for describing a document.
- Decide what metadata fields to present to the user during a search.
- Set up subject hierarchies that provide meaningful browsing options to users.
- Register the repository with OAI. Since OAI is a built-in feature, the registration is easy.

The repository is now ready to accept documents. Authors place documents in a temporary storage buffer. Before moving documents from the buffer to the public area, Caltech librarians perform the following quality-control checks:

- Enforce repository policies with respect to author affiliation, subject area, departmental approval process, or any other criteria appropriate for each repository.
- Verify and (if necessary) improve the metadata. Good metadata enhance discoverability.
- Check document formats:
 - Ensure online readability of all submitted documents.
 - Convert documents to formats that conform to best practices.

- Take one of the following actions:
 - Return the document with comment to the author.
 - Reject and delete the document.
 - Accept the document.
- Create a unique document identifier. We create our own persistent identifiers independently of the Eprints system. This is a safeguard in case we switch from Eprints to another system in the future. As much as we like Eprints now, better systems may come along. Moreover, no one can guarantee the long-term survival of any software. (A detailed description of our identifiers and associated resolver will be self-archived in the caltechLIB repository at <http://library.caltech.edu/digital>.)
- Create browse pages. We generate a browseable view of the repository by executing Eprints-supplied scripts on a regular nightly schedule. These scripts generate static web pages containing subject-grouped lists of links to documents in the repository.

User Features

- Eprints supports any type of document format, including HTML, Adobe PDF, and PostScript. However, repository administrators should carefully consider which formats they are willing to support and maintain.
- Authors must create a repository account in order to be able to submit documents. The repository administrator controls what information is requested and what information can be used to create metadata for submitted documents.
- Readers are encouraged to create an account. Registered users may set up an e-mail alerting service for new content in their subject areas.
- The current version of Eprints assigns each user a password. This is somewhat of an annoyance. Because users cannot choose their passwords, they are likely to forget them.

Overall Evaluation

Eprints is a powerful and inexpensive solution for sharing scholarly works with the world, a concept Harnad calls “self-archiving.” The web-based submission process is intuitive and requires minimal effort on the part of authors. However, long-term preservation requires an institutional commitment. The Caltech Library System is committed to preserve indefinitely those documents that are self-archived in its repositories. To this end, the library performs quality checks on submitted documents and metadata, enforces repository policies, and assigns persistent identifiers. Eprints gives us web-based tools to perform these management tasks efficiently.

Eprints validates many of Harnad’s claims. It is possible for researchers to make their research freely available to everyone, increasing the impact of their research in the process. Because of Ginsparg’s arXiv, physicists already profit from this revolution in scholarly communication. With Eprints and the Open Archives Initiative, the fundamental building blocks are in place to spread this revolution to all other disciplines. The only roadblock is the willingness of researchers to experiment with this new scholarly-communication model.