

Open knowledge: FOSS and Open Access. A case study of ePrints.FRI
Miha Peternel, Faculty of Computer and Information Science,
University of Ljubljana, Slovenia
miha.peternel@fri.uni-lj.si

In the last few years the concept of Open Access to publications has become increasingly important, because academic publishers are trying to restrict access to research publications. In a sense Open Access publication movement is related to Free Open Source Software movement, also there are free open source packages to implement publication databases. The Open Archives Initiative promotes Open Access by standardization of Open Archives software interfaces, automatic indexing of open access publication databases, and educating researchers on the importance and benefits of open archiving for dissemination of research publications.

In the age of internet educational and research institutions need electronic repositories of teaching and research materials. In Slovenia the government requires faculties to publish student theses electronically. Projects funded by European Union also require research publications be published in a freely accessible form. There are several open source packages that can be customized for institutional requirements. At our faculty we have chosen ePrints and customized it to develop ePrints.FRI, the institutional database of Faculty of Computer and Information Science, University of Ljubljana. EPrints is developed at the School of Electronics and Computer Science, University of Southampton, UK. It is open source repository software written mostly in Perl. It integrates with Apache web server and uses a plethora of other open source utilities for backend processing. It can easily be customized by modifying XML templates, plaintext templates, and Perl code.

We started building our archive in 2002 with ePrints version 2. The initial goal was to provide a simple self-archiving tool for our laboratory and other interested researchers at the faculty. Prof. Franc Solina suggested we support Open Archive Initiative and so our archive became the first open archive in a wider region. In 2008 we launched an institutional effort to fulfil national directive on thesis publishing. We revised our policies for digital publishing and decided to upgrade the archive to ePrints 3. We also redefined meta-data to provide a single point of thesis meta-data entry for the faculty information system.

Open source archiving software is only a starting point for an institutional repository. The first phase took a single developer less than 3 months for basic customization, internal testing, dedicated server setup and multilingual debugging. On the other hand setting up and deploying an institutional repository requires a lot of work spanning several departments that don't usually cooperate – in our case there were 4 departments involved so we needed an organized process. We have used a project based approach with one manager, two software engineers, IT department representative, student department representative, library representative, a linguist and institutional representatives. More student office and library personnel were involved in the final phases of the project. All of them except one part-time software script engineer were already employed by the faculty. The developmental phases of the institutional effort were institutional planning, metadata definition, customization, translation, staff education, testing, migration of existing publications from older database, initial deployment, workflow facilitation and statistics.

Besides archiving service implementation, a submission policy was developed. We use different policies for research publications and theses. There are strict requirements for student theses. Students must sign papers submitting rights for electronic publishing. Scientific papers are self-published and checked on best-effort basis.

We modernized institutional workflows to facilitate electronic publication of theses. Students are now required to submit thesis and meta-data in electronic form before oral defense.

Announcement and publication proceed automatically. The publication system has been integrated with other information systems. Faculty web pages include thesis defence announcements and thesis details that are drawn by scripts from ePrints repository, and the content links point to documents hosted in the repository. Also morning mails include thesis defence announcements. Mentoring and committee participation statistics are automatically generated from thesis meta-data.

We have translated most of the web interface to Slovenian language and we had to resolve many language specific issues. Most of the translation was implemented by translating XML language files, however we found missing ePrints language flexibility which we had to work around by modifying both Perl code and XML. Most of the meta-data is bilingual, but OAI interface doesn't standardize multilingual meta-data so we added code to preferably forward English meta-data to OAI indexers.

Faculty committed to the project in December 2007, test installation was in April 2008, institutional deployment for testing was in September and public deployment in October 2008.

Hosting is provided by our IT department on IBM server platform hosting multiple virtual servers under VMWare. ePrints.FRI runs on a virtual Debian Linux server. The backup is provided by IT department. Backup is straightforward as we backup the whole virtual server image, so we can also easily migrate ePrints.FRI to a new server. To sustain service we prepared printed instructions and educated 4 student office and 2 library office staff. The server is administered by one ePrints administrator and one system administrator plus one support engineer. Technical support is provided by IT department and if the problem cannot be resolved it is forwarded to involved technical staff.

Web traffic analysis enables us to monitor interest as well as the origin of interest for publication topics. For general access statistics we use AWStats and Webalizer. Recently we also installed IRStats which is repository specific open source traffic analyzer. It integrates with ePrints database and provides a simple interface to create graphs of repository statistics. We measured publication dissemination to about double the number of enrolled students per month. Most visitors arrive by search engines looking for general keywords.

Open source ePrints software has allowed our faculty to cost-effectively deploy an institutional archive customized to our requirements and workflow, and integrate it with existing web and email information systems. The effects of the Open Access are measured in increased promotion and dissemination of publications through OAI indexes and Google. As a result of electronic publication student theses have been widely disseminated and have thus become a useful reference for graduating students. In the future we would like the software interfaces to achieve automated integration with current and future national archives.