The new EUR institutional repository

By way of introduction

According to Jane Euler from CNRI, Reston Virginia, the Erasmus University Rotterdam was the third DSpace user to register a DSpace handle server; in november 2007 we were the first to deprecate the DSpace handle server and register a new handle server with the same prefix redirecting traffic to the new EUR institutional repository.

I will not discuss our experiences with DSpace in detail. It suffices to say that DSpace was the perfect tool to get an IR up and running quickly. However, after the infamous “Cream of Science” project all Dutch IR’s participated in (2005), we sat back, and overlooking the battlefield, we came to the conclusion that there must be a better, more reliable way to run an IR.

Chop it up

We already had decided that the DSpace interface was not the best way to present the holdings of an IR to the larger public. So we build the RePub website (http://repub.eur.nl/) as the main interface to the IR of the EUR. It works on an OAI-PMH harvest of DSpace, but allows the admins to re-order the DSpace collections into meaningful chunks.

So far, so good. But what about DSpace as a backend for an IR?

The meaning of the word repository

It seems there are basically two interpretations of the word repository in the world of IR’s. One takes the meaning of the word quite literally: One stores items in a repository and that is it. The other sees a repository as an electronic environment to store items to do something with them.

The EUR belongs to the second tribe. We enrich the metadata of publications in order to promote these publications. We correct mistakes if we find them. In short: We work a lot on the metadata of the items in our IR.

The IR provides services for our authors. We have automated uploads to SSRN and RePEc. We have to provide reliable OAI-PMH feeds for services like DAREnet.

We participate in national and international projects that investigate possible services based on the contents of IR’s. Often this implies changes in the metadata in order to accomplish these services (MPEG21/DIDL output, etc.).

We were asked to store datasets and educational material in the IR.

To sum up: We needed an environment that was flexible, could be used reliably by several people at the same time, but above all something that was simple.

The new EUR repository

The new EUR repository consists of two filesystems: One for the metadata, the other for the assets. The metadata filesystem is a Subversion filesystem. The assetstore is a webdav filesystem.

With Subversion we have version controlled metadata. On top of Subversion we run a small Python codebase for business logic and services:
• Postgresql databases for swift access for editors;
• embargo dates, etc.;
• access rules;
• OAI-PMH feeds;
• other services.

Metadata is chopped up into:
• item metadata;
• person metadata;
• organizational metadata;
• sets.

Relations between these metadata, all under version control (and all with date.modified regime), are expressed by way of references to ID’s. When data is used for a particular service the ID’s are resolved.

For workflows we can define so-called “custom-editors” that can be plugged in several ways into Subversion. We use Django, a Python web application framework. We also use Django for the admin interface.

The system went live at the end of November 2007.