Changing Platforms: Parallel case studies in repository platform migration

Overview of case-study

This presentation is based on the authors' personal experience of migrating long-established, fully operational repositories between two of the leading open source, open access repository platforms. In both cases the migrations have achieved successful outcomes. The objective of this paper is to document key decisions and actions of these migration projects, in order to help and support decision-making by repository managers and technical advisers in the process of software migration, or perhaps just considering it.

The authors share the experience of having overseen the migration of their respective Open Access repositories, though in most other respects their situations contrast markedly. One is a major international subject repository, the other an unprepossessing IR in a UK research school. One has many tens of thousands of items, the other barely 2000 items. One has a distributed international editorial network, the other a single local repository manager. They are clearly at different places along the range of endeavour in the Open Access repository community. And for that reason, a combined examination of both their experiences ensures a good spread of activities and outcomes to analyse and share with the wider community.

In 2010 E-LIS[1] migrated from EPrints to DSpace. E-LIS is an international Open Archive for Library and Information Science (LIS). E-LIS used Eprints Software since 2003 starting with version 2.3, upgrading over time to version 3.1. E-LIS is a freely accessible subject repository, and a voluntary enterprise. The E-LIS team reported on their experience upgrading Eprints software from version 2 to 3 at Open Repositories

[1]E-LIS http://eprints.rclis.org/ Last accessed February 2011

2009[2]. The SAS-Space repository[3] was set up as an Institutional Repository for the University of London's School of Advanced Study in 2006, using DSpace version 1.5. An account of its creation was given at the DSUG[4] in Rome in 2007.

For each of our separate projects, we describe the key factors that led to the decision to migrate, the process undertaken, and some of the outcomes. A comparative approach allows us to explore several dimensions of potential relevance to a range of repository endeavours; we consider whether platform migration should be considered an extreme and undesirable activity, or whether it is a normal and unremarkable activity, sometimes necessary to maintain the integrity of repository in the continually-evolving repository landscape.

What makes people change platforms?

Nowadays many institutions operate repository systems, and many more are thinking about installing one. What should they consider before any action? A great deal of thought often goes into choosing a repository platform in the first place. Typical considerations include costs, features, and available skills and infrastructure.

Resources like the *Repository Software Survey by the JISC Repositories Support Project*[5] are invaluable. No less valuable are the recommendations of colleagues and other professional contacts in the field: another comparable institution's choice of repository software is always worth considering.

But is it possible to "get it wrong", and, if so, is that necessarily a bad thing? What if circumstances change? Costs of maintaining a particular platform may increase. Key skills, essential to maintain, upgrade, customise and configure the repository may no longer be available, and therefore it is necessary to expand the range of opportunities to hire technical specialists for repository management.

New and necessary features, may become available in other platforms, with little prospect of them being developed for one's own platform in an acceptable space of time. For example: implementations of new standards for metadata; workflow tools;

---

[2]Imma Subirats, Antonella De Robbio, Zeno Tajoli. "Library and Information Science Open Access: a review of the last six years in an international multilingual environment" at the *Open Repositories 2009* http://eprints.rclis.org/handle/10760/13123 Last accessed February 2011

[3]SAS-Space repository http://sas-space.sas.ac.uk/ Last accessed February 2011

[4]Richard Davis. "SAS-Space: a DSpace repository for the School of Advanced Study, University of London" at the *DSUG 2007* http://www.aepic.it/conf/DSUG2007/viewabstract8587.html?id=331&cf=11 Last accessed February 2011

[5]http://www.rsp.ac.uk/start/software-survey/results-2010/

document management features; end user services; administrative tools such as statistics; new export/import/deposit features.

Other legitimate imperatives include minimising the cost of software maintenance, or strengthening the network of institutional contacts using the same software, for mutual support and assistance.

Should repository managers accept that they are tied to a particular platform - be fiercely, even blindly loyal to it and “soldier on”? Should they consider the choice of platform irrevocable and permanent? Does the need to change platforms indicate some kind of failure in the system implementation; or is it all right to change if institutional needs also change?

It is our belief that the real failure would be to not do what is necessary to keep the repository alive, viable and effective, within its particular and unique institutional context. And if, after full and frank consideration of the costs, benefits and consequences, that may legitimately necessitate a change of platform.

Key points of presentation

The motivation for the two migrations was very similar: in both cases it was dictated by the availability of skills and resources. In the case of E-LIS, specialists willing to voluntarily host and support the repository were only available from the DSpace community. In the case of SAS-Space the in-house development team and infrastructure favoured EPrints. (A few blessed institutions have repository specialists fluent in more than one repository package, however this is still rare.)

EPrints and DSpace are often contrasted with each other: their differences are often defined by their implementation (Apache/Perl in the case of EPrints; Tomcat/Java in the case of DSpace). Nevertheless they have many features in common. This is hardly surprising since both packages share a common lineage in earlier digital object management initiatives, notably Cogprints, both share a common adherence to the same open standards and open access, and there are even developers who contributed to both.

One benefit of this common heritage is that exporting and importing between the two packages was relatively easy, or at least easily comprehensible: thanks to the similarities between their approach to arranging exported digital object bitstreams and metadata, the transformations necessary were relatively straightforward. Nevertheless a

considerable amount of testing and checking was necessary; and we will highlight a few traps for the unwary.

Some critical differences exist nevertheless. For example, it was also necessary to consider the implications of losing (or gaining) a Handle based URI system; of implementing a different approaches to representing institutional structures and intellectual classifications; of mapping different object type definitions; and of consolidating different approaches to gathering statistics. In addition both projects also had to weigh the implications of an earlier decision to embed references to their software platforms in their chosen domain names.

Implementation and synchronisation of the new systems also needed careful preparation and execution. Both projects had to ensure as little disruption and as smooth a transition as possible for their users. This involved being aware of the impact on other dynamically linked systems. In the case of E-LIS, too, there was a large multinational group of editors and users to consider, all with a longstanding familiarity and attachment to the outgoing interface.

There may still be reverberations of the migrations as yet unnoticed. In both cases, however, the outcomes appear to have been a success, and met with an encouraging reception from repository managers, editors and users. The fact that substantial collections of metadata can be transferred successfully between systems should be seen as positive for the community; and one conclusion of our reflections is that the ability to migrate large digital objects collections, and their metadata, reliably and effectively between software platforms, is an important freedom, made possible only by the use and availability of common open standards and tools.