

Specifying Repository Requirements

support@rsp.ac.uk

Overview

Specifying a repository system involves technical choices that should be led by a general requirements analysis: what is the repository required to do, and for whom? This briefing paper shows how to identify those broad requirements that can inform a system specification.

Preliminaries: putting technical issues aside (for now)

A systems requirements analysis is an essential pre-requisite to choosing the right software to build a repository. Repository software choices span a range of costs, support different needs for content diversification, customisation and quality, and serve a range of stakeholders.

The analysis that leads to the specification is not predominantly technical, and should involve all members of the repository team, not just systems staff, even though systems requirements may ultimately be expressed in technical terms. First, what is the repository for? Who are the stakeholders -those people with a vested interest in how the repository represents the institution, and themselves, to the world -and what do they want from the repository? In the case of an institutional repository, stakeholders will include senior institutional managers, departmental leaders, and those who are expected to contribute content. This approach is likely to reveal a series of questions:

- What is the target content of the repository?
- Are all content types to be managed in a single repository, or more than one?
- What other systems and services might the repository be required to share information with? This is often referred to as 'interoperability'.
- Does the budget support the requirements?

Content management

For a higher education institution, repository content could include research papers and data, electronic theses, creative arts research, as well as teaching and learning resources, including some audio-visual content.

As repository software has become more flexible in handling a wider range of data types, the type of information is less of a factor in specifying systems requirements.

Interoperability

For repositories, efficiency and flexibility in getting data in, and out, in a variety of formats, are key requirements. Interoperability of institutional repositories has largely been founded on the Open Archives Initiative (OAI), which first defined how repositories as data providers could collectively become visible and searchable through OAI service providers. Now interoperability is also likely to embrace OAI-Object Reuse and Exchange (ORE). content deposit protocols such as SWORD, as well as web services standards, including Web 2.0, digital library systems and other institutional systems, as well as personal information systems such as citation applications.

Perhaps the most important aspect at the moment, is the need to integrate with internal research management systems which can be commercial or bespoke. Commercial packages such as Symplectic, Pure and Converis have developed tools for interoperability with the main repository softwares.

Standards can enforce interoperability. OAI, Dublin Core, and W3C accessibility are core standards for research repositories. Greater interoperability with services can provide flexibility for upgrading the repository in future.

Data models such as CERIF (Common European Research Information Format) are being adopted by

HEIs to allow exchange of information with other CERIF compliant systems.

Keeping within budget

The most obvious consideration governing a system specification is the budget: how much money is available to support the system? Repositories are a long-term investment for the institution. Establishing a working repository that is easy to use for the purpose intended must be the first target if it is to have an impact and develop working practices and cultures. Almost every decision will be influenced by cost. Be realistic, but remember that a requirements analysis based on a full stakeholder consultation will result in a stronger position to secure the funds needed to fulfill the needs the stakeholders have identified.

Open source software, hosted software or repository services?

A key factor in cost considerations is not which software but how it is delivered and supported. Open source software (OSS) is free to download, install and use, but there is an ongoing cost to implement and maintain the repository, depending on the complexity or ease of use of the software chosen and the ability to install and configure it for use. Good support services and documentation can reduce implementation costs. OSS does not typically promise to support end-users directly, but support lists and communities can often assist. Hosted software providers, on the other hand, often charge a fee to provide a supported software environment. Repository services are an extra choice today. Users can pay for specified services such as repository

building, hosting, customisation and optimisation, rather than software. In this way they could combine OSS with paid support to improve cost efficiency of the repository.

Some technical considerations

Decisions about software and support may rest on the systems support available locally. That support may, in turn, influence the next requirements, particularly:

- Platform requirements: Which server operating system? Typically, OSS repositories use other open source components, such as a web server, and a database. Linux has become the operating system of choice, but some repository OSSs now offer a Windows version as an alternative.
- Programming requirements: Does the repository need programming skills? Programming helps develop, customise and extend the repository. Unless significant development of the repository is anticipated, basing choice of repository on this criterion could be ceding other important requirements to a single technical decision. In the latest versions of some repository software, graphical user interfaces reduce the programming requirements, giving repository managers and content administrators greater control.

Conclusion

This briefing paper has shown how to identify some of the key issues and priorities affecting repository requirements. It has not shown how to expand that analysis into a full systems requirements analysis, nor what software might fulfill the requirements. To do that, see the Repository Software Options on the RSP Web site.

References & Further information

References & further information:

Simple Web-service Offering Repository Deposit, SWORD <http://www.swordapp.org>

Open Archives Initiative Object Reuse and Exchange, OAI-ORE <http://www.openarchives.org/ore/>

CERIF <http://bit.ly/rQJ9sy>

Commercial Repository Solutions, RSP technical briefing paper <http://bit.ly/ubzMRX>

RSP Repository Software Options <http://bit.ly/ttafa0>