Universities of Leeds, Sheffield and York
http://eprints.whiterose.ac.uk/

# White Rose Research Online: A Case History

February 2008

*Rachel Proudfoot / White Rose Research Online Steering Group*

## Contents



## 1. Brief outline

- White Rose Research Online is a shared "institutional repository" for the Universities of Leeds, Sheffield and York.
- Created in 2004 as part of the national SHERPA project.
- Runs on EPrints software (3.0.2-beta-3 at the time of writing) – server located at University of Leeds
- Holds predominantly textual research outputs.
- Items: 2730 open access full text: 93%
- Average stats per month:
  - Google Analytics: 7860 visits / 19,860 pageviews
- Staff (shared by the three institutions)
  - 1.4 FTE permanent staff (Repository Officer; Repository Assistant)
  - 1.8 FTE project staff (Project Manager, Project Officer; Software Developer (from March 08))
  - Ad hoc systems support and 0.1 FTE clerical support

Some statistics:

| | Leeds | Sheffield | York |
|---|---|---|---|
| Degree Students (FT&PT) | 30,494 | 23,914 | 11,820 |
| Staff | 7,993 | 5,951 | 3,080 |
| Academic & Research staff | 2,457 | 2,377 | 1,210 |

## 2. Why a consortium repository?

The three White Rose universities have collaborated in a number of areas over the past few years e.g. the White Rose Centre for Excellence in Teaching and Learning in Enterprise; co-supervision of PhD students; White Rose high performance computing grid etc. The three White Rose libraries also have a cordial and collaborative relationship; the three library Directors meet each term.

Some of the key reasons the libraries cooperated in creating a shared repository were:

- The libraries' history of cooperation
- Endorsement of White Rose collaboration at senior levels across the three institutions
- White Rose Consortium being seen as a potential source of funding for the repository (this proved not to be the case as WRC does not have grant money)
- White Rose "brand": a collaborative repository might prove attractive to University management and individual researchers
- Pooled resources would allow the libraries to employ a FT, shared repository officer
- Possible economies of scale (such as single server and software instance)
- Partnership would offer a good opportunity to share expertise and experience
- The repository could have a regional role in helping researchers identify local collaborators. (To the best of our knowledge, the system has not done this to date though, perhaps not unexpectedly, an analysis of recent research outputs from the three institutions suggests an increasing level of White Rose co-authorship.)
- To explore the practicalities of a consortial repository; possibly providing a model for other regional consortia

There are both pros and cons to this type of consortial approach but the repository is likely to remain a shared system for some time to come – with one key driver being the government's current emphasis on shared services.

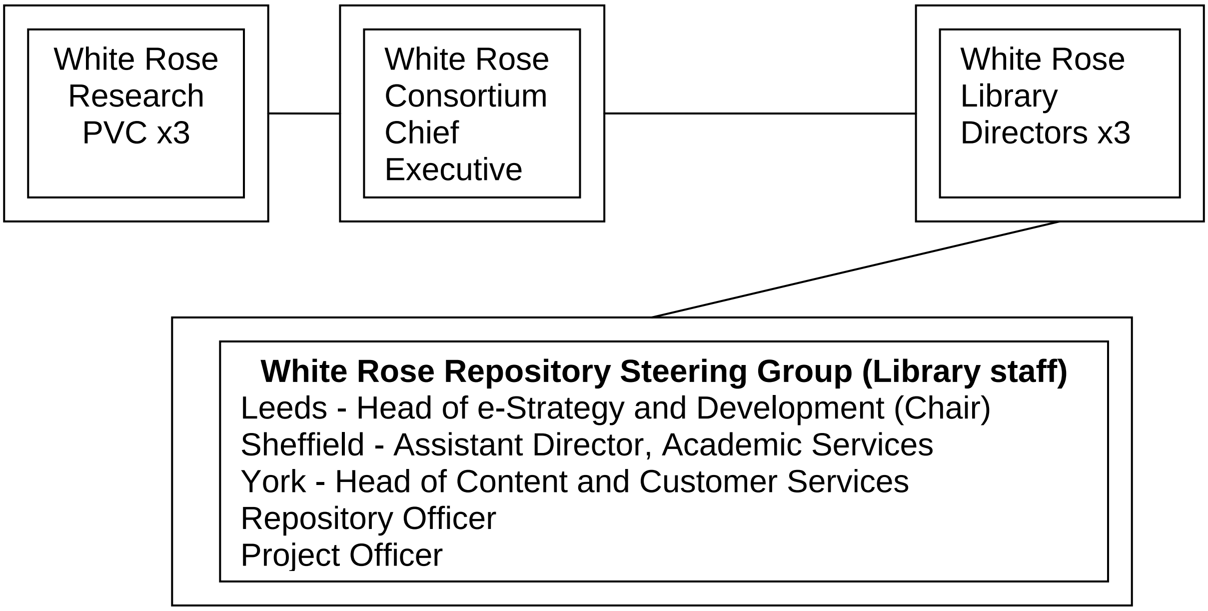
## 3. Governance

For the first three years of its life, the repository has been shared largely on the basis of goodwill between the partners rather than on a more formal footing. As the small scale project repository starts to transform to a sustainable service for the three institutions, more formal agreements are being put in place. So far, there has been regular discussion and communication about repository development and significant differences of opinion between the partners have been avoided. Inevitably, there are some political aspects to repository development: in particular, it has been important that no one partner should dominate the repository either in terms of decision making or content.

The initiative has been library led with some input from academic staff towards the beginning of the SHERPA project and subsequently via committee feedback. The WRRO Steering Group also has a link to the three White Rose Research PVCs (who meet regularly) to report progress and seek guidance as required.

However, moves are afoot to increase researcher input into shaping repository policy.

**Organisation Chart**




## 4. Collection policies

- The initial collection policy was to target peer-reviewed journal papers and to offer 100% open access full text. The Repository Officer actively identified and sought “green” content – particularly if the published PDF could be archived. The narrow collection focus of the repository was informed by decision made by the Repository Steering Group, feedback from research committees and a repository working group at the University of York.
- Subsequently, this policy expanded to the collection of “RAE type” research outputs, including conference papers, a range of journal papers and book chapters. Repository staff started to contact researchers to ask for new outputs recently indexed in Web of Science; this activity is still underway. The aim is to increase timely capture of files and provide ongoing publicity for the repository – though the researcher response rate is modest.
- The most recent phase sees a more embracing approach to collection; something which is hoped will be attractive to depositors. We will collect any published research outputs and make as many outputs as possible openly available. Inevitably, some works will be embargoed and there may be others where the researcher does not have the relevant version of the work or where we do not have permission to make the work openly available. Increasingly, we are working with relevant administrators in departments and faculties who provide a link to - and support for - research staff.

A key question for institutional repositories – and one we are still grappling with – is whether to strive for a comprehensive listing of research outputs, even if this means a high proportion of metadata-only entries. A related question is whether the repository is primarily a tool for research dissemination or whether the repository has a role to play in research administration.

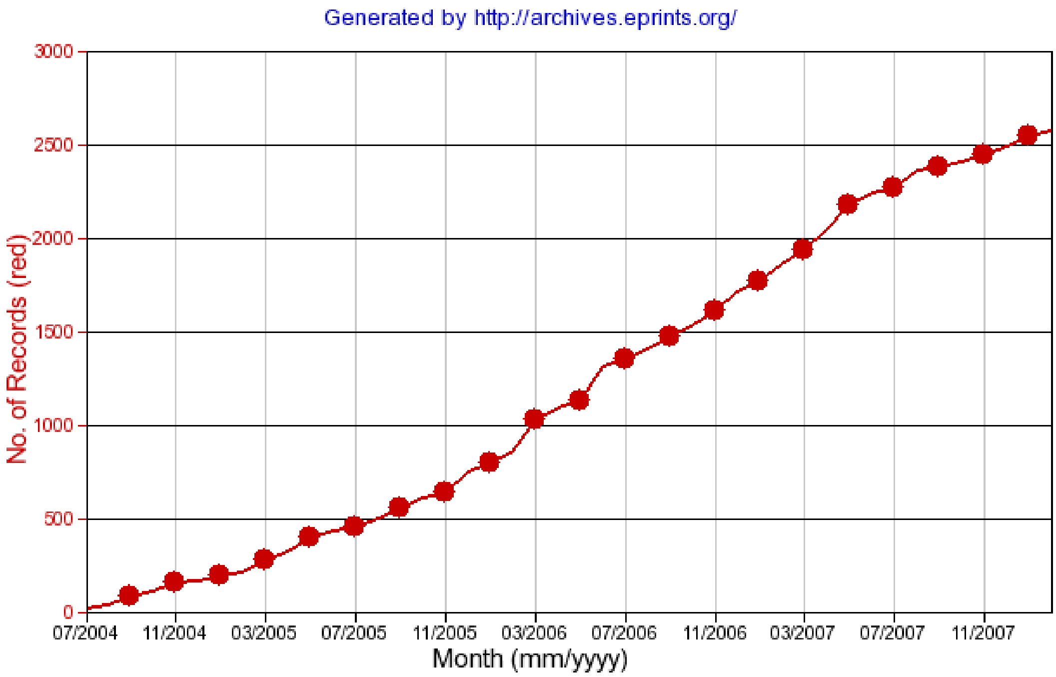
## 5. Advocacy activities

In common with many other repository services, we have undertaken a broad range of advocacy activities over the past three years or so, including:

- papers to senior University committees (Library Committees, Research Committees / Board)
- information on the web: repository FAQs and news, library web sites
- presentations at a variety of meetings: subject librarians, library reps, departmental committees
- handouts and displays about open access
- emails to academic staff e.g. publicity email to Heads of Department
- articles in a variety of local online and paper newsletters
- provision of access and download statistics (to individuals and in departmental reports)

Awareness of WRRO has been and continues to be a significant issue. This echoes the experience of other institutional repositories but the challenge of raising awareness is perhaps more complex for a consortium.  Certainly, the scale of the task is significant and though we have undertake a number of activities, findings from a recent questionnaire suggest about one third of researchers have heard of White Rose Research Online – so there is quite a lot of advocacy work still to do. Awareness is highest amongst senior academics.

## 6. Growth & the IncReASe project



Growth has been steady – but modest – reflecting only a small proportion of research outputs at the three institutions. Matched funding for the IncReASe Project (*Increasing Repository Content Through Automation and Services*) has been secured from JISC and it is hoped the project will see WRRO grow significantly. IncReASe addresses some of the technical issues we have identified over the past three years that are holding the repository back and aims to increase usage of the repository by demonstrating its usefulness as a service – for example, how researchers can be helped to meet research funder open access requirements.

It is estimated the three Universities produce 9-10,000 research outputs per year (taking a relatively narrow, print focussed definition of “research outputs”).

**7. Scalable and sustainable deposit models**

We are anticipating that deposit will continue to be a mixture of:
i) some centrally supported archiving
ii) direct archiving by researchers
iii) proxy archiving by a variety of support staff in faculties / departments

It is anticipated that model (iii) will become increasingly important in creating a scalable and sustainable service. We have some good examples of departmentally based archiving by research support staff and information officers who can liaise closely with the relevant academics and receive support from central repository staff.

**8. Locating the repository in the institutional, national/international landscape**

The repository has developed as a stand alone system to handle textual research outputs. Two of the three partners are creating additional repositories to house a range of multimedia outputs – for learning, teaching and research. Development work is needed to align WRRO with other IT systems across the three universities. Some areas for attention are:

- authentication
- repository access via a university portal (where available)
- dynamic linkage with publication databases and research management systems
- a logical (local) repositories infrastructure

Researchers and managers are often keen to know what similar activity is taking place elsewhere. It is a good idea to have examples from comparator institutions to hand – these can provide reassurance, be a driver for action, demonstrate local success.

**9. Unresolved issues / future opportunities**

Some of the areas where a need for action has been identified and/or where development would improve the repository service are outlined below.

**Community engagement:** more opportunities for researchers to inform the development of the repository.
**Systems**: more robust storage and backup; aligning EPrints with local University systems e.g. looking at Oracle; interfacing with local research management systems; introducing authentication (Shibboleth).
**Preservation:** something all repositories will need to tackle; we will consider what good in-house practice we can adopt and hope for the development of affordable external preservation services.
**Services:** easy/automated content import/export to/from other internal and external systems.
**Statistics:** should be available for individual papers, by author, by department / faculty; a number of readily generated reports are needed to demonstrate usage profiles and search routes into the repository.
**Metadata enhancement / consistency**: descriptive, preservation and administrative. E.g. Use of authority lists to improve metadata consistency.
**Rights:** exposure of machine readable licences to maximise reuse opportunities where possible.
**Data input:** we need to make the most of opportunities to import pre-existing metadata and to work with administrative staff across the consortium to create sustainable, scalable ingest processes. Individual self-archiving will be encouraged but we do not anticipate it will be the prevalent archiving model.

## 10. Lessons learned: summary

- There are some economies of scale in a shared repository system; it also raises challenges when looking at interfacing with local IT systems. It is challenging to target such a large and diverse body of researchers.
- Think big, but be realistic about what can be tackled most effectively with limited resources.
- Think carefully about the rationale for your system and how to balance the “top down” managerial demands of the system with the “bottom up” research dissemination demands.
- A consortial approach can be attractive to senior managers; individual researchers seem to be neutral about the consortium arrangement – to date, the shared nature of the system has not been a strong incentive or disincentive to deposition.
- Most publishers have accepted the consortial repository as equivalent to a single “institutional” repository; but not all. The question of whether the repository forms part of the “employer’s web site” becomes more ambiguous.
- Researcher awareness of funder open access requirements is still low (Feb 08).
- We need to understand more about disciplinary differences in research workflows and outputs and consider how to reflect these in repository collection policies, ingest processes and outputs.
- Usage statistics are probably the biggest selling point for researchers.
- Don’t underestimate the level of technical support required for both support and development.

- It is difficult to recruit technical staff: consider grading, advertising and appropriate job titles to reach your target group ("institutional repository" may not mean very much to software developers).
- A purist "open access only" approach may be difficult to maintain; is it better to have a comprehensive collection policy with the facility for restricted access where necessary?
- Any software upgrade or major local customisation should be followed by test searches across various engines and harvesters to check metadata is still being exposed correctly.
- Attitudes towards open source software at the institutional level can be negative.
- Sharing experience with other repository colleagues is invaluable.

## 11. Relevant links

White Rose Research Online
http://eprints.whiterose.ac.uk/

IncReASe Project Web Site
http://eprints.whiterose.ac.uk/increase/

Proudfoot, R. (2005) *The White Rose Consortium ePrints Repository: creating a shared institutional repository for the Universities of Leeds, Sheffield and York.* ALISS Quarterly, 1 (1). pp. 19-23
http://eprints.whiterose.ac.uk/858/

White Rose Consortium Web Site
http://www.whiterose.ac.uk/