



**Progress Report Template -**

*Text in italics is explanatory and should be deleted in completed documents.*

<b>Project Name</b>	Increasing repository content through automation and services, University of Leeds
<b>Project Website</b>	<i><a href="http://eprints.whiterose.ac.uk/increase/">http://eprints.whiterose.ac.uk/increase/</a></i>
<b>Report compiled by</b>	<i>Rachel Proudfoot</i>
<b>Reporting period</b>	18/10/07 – 25/04/08

**Section One: Summary**

During the last reporting period we have completed our web site survey across the Consortium and have administered an online questionnaire to researchers. The survey found scattered full text across the three web sites – predominantly on individual web pages – but little systematic handling of either metadata (with the exception of the University of Leeds publication database) or full texts. We were hoping to find databases which could be imported into WRRO but these are few in number – particularly when narrowed down to organised collections of full text. We plan to work initially with the Computer Science department at the University of Sheffield to scope importing their publication database and enhancing this with as much full text as possible. The online researcher questionnaire revealed – not entirely unexpectedly – that awareness of WRRO amongst research staff is nowhere near what we would like. The questionnaire included a variant of the well known Swan and Brown (2005)<sup>1</sup> question on attitudes towards open access mandates: our findings more or less replicated their results – though with a higher level of “reluctant” compliance. Our findings showed that awareness of funder mandates amongst researchers is still quite low. We also identified a number of potential drivers for deposit. The questionnaire has proved useful publicity for the repository, resulting in a small flurry of new self-archivers.

The recruitment to the full time Software Developer post proved difficult with the post being filled from March 08, eight months into the project. During the interim, the project has continued to receive some technical support from within the University of Leeds Library Systems Team. Induction is well underway for the IncReASe Software Developer and we are making progress on technical aspects of the project. We have a rough plan of action for work with ESRC. ESRC is keen to move forward in investigating how material can be deposited in one place in order to feed both an institutional and funder repository. We have suggested various scenarios, the bare bones of which are outlined at: <http://eprints.whiterose.ac.uk/increase/esrc.html>. ESRC grant holders will be contacted during the interview phase of IncReASe.

**Section Two: Activities and Progress**

The overall approach to the project has changed somewhat in the light of staff availability. In particular, liaison with pilot departments which have importable databases is still ongoing and likely to be so for some time; any bulk import of metadata from departments can only take place

---

<sup>1</sup> Swan, A. and Brown, S. (2005) *Open access self archiving: an author study*. Cornwall: Key Perspectives. A report undertaken for JISC and available from [http://www.keyperspectives.co.uk/openaccessarchive/reports/Open%20Access%20II%20\(author%20survey%20on%20self%20archiving\)%202005.pdf](http://www.keyperspectives.co.uk/openaccessarchive/reports/Open%20Access%20II%20(author%20survey%20on%20self%20archiving)%202005.pdf)

now that the Software Developer is in post.

#### WP1 Project Management

- Consortium agreement submitted to JISC.
- Update reports for internal use across the consortium have been circulated – to keep library management and other interested staff up to date with progress.
- Attendance at JISC Programme Meeting.
- Induction for the *Software Developer: IncReASe Project* well under way.
- Revision of milestones document.
- Agreed to web site archiving with UK Web Archiving Consortium (UKWAC).

#### WP2 Survey of activities

- Web site analysis complete and results online.
- *Database Prevalence Report* complete. (Milestone 10)
- Interesting case studies identified for follow up.
- Online researcher questionnaire designed and administered; analysis underway. Preliminary findings and text of questionnaire online. The questionnaire request was sent by the three White Rose PVCs for Research. This resulted in a high initial response which tailed off quite quickly. Overall we received 330 responses, with a half or dozen or so disregarded

We have not achieved our goal to import three pilot databases (milestone 3) nor written the Pilot Test Report (milestone 4). We will complete the relevant activities and milestones within the lifetime of the project but they will be delivered later than anticipated.

#### WP3 Harvesting

- Introduction of additional notes field in Eprints repository buffer – allows for better communication across several different editors.
- We undertook field mapping against the University of Leeds Publication Database and agreed on a deposit workflow for academics; however, this work is on ice because of the University's publication data review. We should know Leeds' plans in the next few weeks; these may point to a different way forward for embedding WRRO in Leeds systems.
- The web site survey revealed the extent to which bibliographic information is inconsistent both within and between web pages. For this reason, we do not think that a web scraper PERL script will have a broad application for auto-harvesting of publication metadata from authors' web pages. Nonetheless, we can see scenarios – where an individual has a large list of publications with a high degree of consistency – where a scraper script could be attractive. The scripting is underway and we anticipate availability within the next few weeks.

#### WP4 Metadata enhancement

- We will aim to comply with the research funder guidelines for grant acknowledgement as described in the RIN guidance document at <http://www.rin.ac.uk/funders-acknowledgement> We are considering whether we need three field for grant data: funder(s) name; grant number(s); grant acknowledgement(s).
- Review of scholarly works application profile is nearly complete. We have thought about the implications of full implementation of SWAP for our local repository. However, having seen Eprints 3.1 demonstrated at the Southampton OR08 Conference, we can see that this would enable us to assign metadata at the document level. An upgrade to 3.1, once available, may be a more effective way to become more SWAP-like than enhancement work done locally.

#### WP5 Integration with funder repositories

- We have proposed four deposit scenarios to the ESRC: two assuming initial deposit in the institutional repository and two assuming initial deposit in the ESRC's Awards and Outputs repository. One way to populate both repositories from a single deposit would be to use OAI-PMH harvesting; another way would be to put data deposited in one repository into the other repository using SWORD. It is likely both methods will be investigated. A summary of the proposed scenarios is online at: <http://eprints.whiterose.ac.uk/increase/funders.html> We are very keen to receive any feedback or suggestions from anyone in the repository community about their views on the optimum method for deposit or other issues we should consider when discussing possibilities with ESRC.
- ESRC fundees have been identified at the University of Leeds and we have started to explore sources of grant metadata and how this may be made available to WRRO. Preliminary discussions have also taken place with York but we have not yet identified a suitable source of grant information.
- ESRC fundees will be included in our interview phase which may help inform our work on feeding the local and ESRC repository.
- Data on repository content and usage statistics was provided to some subject librarians, on request, for inclusion in their annual reports to departments. It is clear, though, that we need to improve both the way statistics are generated and the facilities for subject librarians to generate relevant reports as required. At the moment, we are using Google Analytics. This is useful at the top level of the repository to show the overall profile of usage, who is accessing the system, the routes into the system and so on. However, it is of little use at the departmental and individual author level. As an interim measure, download statistics are being generated directly from our access logs.

#### WP6 Integration with local systems

- A workflow was agreed with the University of Leeds Publication Database but this action is on ice pending University level decisions about how research outputs will be handled.

#### WP 7 Evaluation and dissemination

See *Evaluation* section

### **Section Three: Institutional & Project Partner Issues**

The IncReASe project and the University of York SAFIR project share a member of staff (based at University of York). This is potentially beneficial in bringing together common interests and issues. Both York and Leeds have more than one repository system for housing different types of digital content (York Fedora and Leeds Digitool). Discussion is ongoing about how these different systems fit together. It might be useful to share experience with other multiple repository institutions. In particular, it will be important to ensure that deposit processes for research and teaching materials are not overly complex.

As mentioned elsewhere in the report, a research management review is underway at each of the three partner institutions including consideration of how publication data will be managed for REF. This will certainly impact WRRO and may also influence some aspects of the project - in particular, WP6 looking at integration with local systems. The research review introduces some new risks to the project but also offers positive opportunities to become more embedded in any emerging systems.

Each partner is looking at creating a local advisory group, including senior academics who may be able to influence decision making at Research Committee/Board level. It is hoped this will be useful in driving forward both the project and the WRRO service.

#### **Section Four: Outputs and Deliverables**

A detailed survey of the web sites from the three White Rose partners was undertaken, with a view to filling in our picture of metadata and text availability across the Consortium and, in particular, to locate where organised databases of metadata – and possibly full text – are located in the institutions. A summary of our findings forms our **milestone 10** output, the **Database Prevalence Report**.

Online at: <http://eprints.whiterose.ac.uk/increase/texts&databases.html>

We administered an online questionnaire to researchers during February and March 08. The main aims were to gauge awareness of and attitudes towards funder mandates, gauge awareness of WRRO, ask where researchers deposit their work (inside and outside University networks) and also to gain some publicity for WRRO. The questionnaire and a preliminary summary of findings is online at: [http://eprints.whiterose.ac.uk/increase/quest\\_summary.html](http://eprints.whiterose.ac.uk/increase/quest_summary.html)

More detailed analysis will be online soon.

We have identified a number of areas for discussion with researchers in one-to-one / small group interviews. Interviews have started but it is early days so it is possible we will amend these areas. We are interested in comments about the appropriateness of the discussion areas; would it be valuable to include other topics? Our summary is here:

<http://eprints.whiterose.ac.uk/increase/interviews.html>

A Case History of White Rose Research Online was produced for the Open Repositories Conference in Southampton and is online at:

<http://www.rsp.ac.uk/repos/casestudies/pdfs/leedsyorksheffield.pdf>

Our work with ESRC is at a very early stage but we would like comment on our tentative deposit scenarios. These are summarised here:

<http://eprints.whiterose.ac.uk/increase/esrc.html>

#### **Section Five: Outcomes and Lessons Learned**

We struggled to recruit to our technical post. We changed the job title in the hope of attracting more applicants:

- *IncReASe Technical Project Officer* = 1 applicant;
- *IncReASe Project Technical Developer* = 2 applicants (one was an excellent candidate but turned down our offer)
- *Software Developer: IncReASe Project* = 15 applicants. However, this advert attracted a lot of standard software developer CVs and very few applicants addressed themselves to the project or showed any awareness of repositories, open access and so forth.

*Lessons learned:* don't include your project name in a job title (at least, not as the first part of the title); using a more generic job title will attract a wider pool of applicants; the down side of this is the proportion of applicants exhibiting little awareness of your particular requirements; it can be difficult to recruit technical staff – have a good contingency plan.

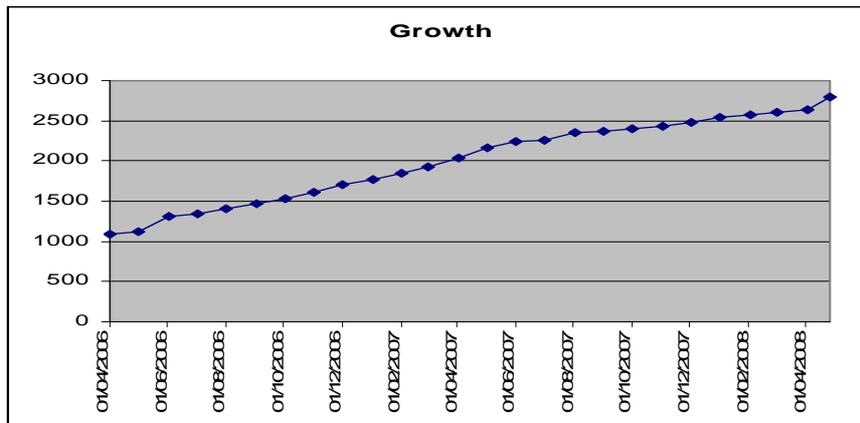
*Lesson learned:* publicity needs to be constant to penetrate researchers' awareness –

particularly in a consortium service. The modest levels of awareness at the partner institutions highlighted in the online questionnaire – despite publicity for quite some time now – underlines this point.

*Lesson learned:* grab researchers' attention by using the title of their most recent publication in the email subject line.

### Section Six: Evaluation

We continue to monitor growth with a view to doubling growth rate by the end of the project. This has not happened so far but the impact of having a full team in place is starting to be seen. Below is the growth for the last two year showing the increased rate for April 08. We hope to improve on this: we need to make a very large jump in ingest rate if we are to capture anything like 20% of new research outputs. It's possible that local research management developments could have a beneficial effect on ingest rate.



We are currently reviewing our communication strategy for WRRO and hope to push information to key stakeholders on a more regular basis with a view to increasing awareness of the repository: routes include the WRRO blog, the White Rose JISMAIL list for library staff, a White Rose JISMAIL list for researchers (though we have not persuaded many to sign up!), the usual local web and paper publicity channels.

### Section Seven: Dissemination

Stakeholder	Engagement
White Rose research staff	Questionnaire – including info to PVCs Schedule of one to one or small group interviews Etheses mandate in Sheffield
JISC Repositories and Preservation Programme	Attendance at programme meeting and contribution to jiscmail list
Repository community within the UK	RSP event hosted by University of Leeds including a presentation on <a href="#">Funder Mandates</a> by Rachel Proudfoot
Global repository community	Case study to OR08. Meeting with Japanese representatives from Hiroshima University Library and Chiba University Library (interested in consortium repositories)

Research Councils	Undertaking scoping exercise with ESRC
Research support staff within institutions	Awareness session for Environment faculty administrators, University of Leeds

### **Section Eight: Risks, Issues and Challenges**

A number of institutions are likely to review their research management systems in the light of REF requirements – this is happening at all three White Rose institutions and presumably at many others. There is an opportunity here for library and repository staff to become involve with the research management review process and the design and implementation of any new or enhanced systems. We are anticipating that local developments may well have an impact on the project because there may be changes in researcher workflows and in the possibility for the collection of both metadata and full text. At least one system currently on the market emphasises its ability to harvest pre-existing metadata – possibly from openly accessible web sites but also from commercial bibliographic databases. There are inevitably some licensing issues about extraction and reuse of metadata but the promise of “crawling the web” to ingest publication metadata – and possibly also gather citation intelligence - is a beguiling one for University decision makers.

As well as positive opportunities, the research management review introduces new risks:

- non-inclusion in the review could lead to WRRO being isolated and make it more difficult to link effectively with any new system;
- the institutional emphasis may become skewed towards research administration rather than dissemination
- any new system may be perceived as a “rival” to WRRO by potential depositors.

We need to make best efforts in our local institutions to inform the review process and emphasise the relevance of WRRO to REF and citation impact.

The need for us to tackle author identification and authentication poses some tricky challenges in a consortium. Nobody involved has experience of, for example, authentication via multiple LDAP servers. Nonetheless, it would clearly be beneficial for users to be able to use their local login credentials; the services we can offer – such as publication lists and tailored statistics – will be limited if we are not able to identify authors unambiguously and link authors with their own publications.

A number of the project milestones which should have been attained a few months into the project have been moved towards the end of the project. This means that a larger number of reports will be clustered in the project’s last couple of months. Although not ideal, there is the advantage that reports at that stage should benefit from drawing on the whole project rather than its initial stages alone.

### **Section Nine: Collaboration and Support**

*What areas of work would you like to discuss with other projects?*

We’d be interested in any feedback about deposit into ESRC’s repository. ESRC have not had many enquiries from institutional repositories so it’s not clear what the demand for a deposit mechanism would be. Are other repositories assuming researchers will deposit locally and ESRC will harvest the works? What advice, if any, are other repositories offering ESRC funded staff?

Are any other projects looking at SWAP? Are they planning to take any action to modify their repositories?

*Is there anything that you would like advice and support on? Do you have any specific training needs, requests or suggestions for supporting workshops for the programme?*

Same as last time: testing SWORD in the context of UKPMC deposit.

It would be useful to have a briefing paper about SWORD for use with university IT staff and decision makers.

We have a new member of staff from a research background who notes that little of the OA debate in the scientific literature discusses institutional repositories. Coverage tends to be skewed heavily towards open access publishing. How do we make inroads into a range of publication read by researchers?

#### **Section Ten: Financial Statement**

*See Appendix A*

#### **Section Eleven: Next Steps**

- Import of pilot database(s).
- One to one / small group interviews with researchers. Hopefully also with administrators.
- Possible upgrade to EPrints 3.1 (depends on release date and stability)
- Demonstrator web scraper PERL script and brief report.
- Meeting with ESRC and software developers to agree the on or more repository feeder mechanisms to be tested and test agreed mechanisms with sample data.
- Agree format for funder metadata.

## Appendix A: Financial Statement

<b>Total Grant</b>	£91,489	<b>Duration of project</b>	Extended from 18 to 20 months
<b>Reporting Period</b>	October 2007 – April 2008		

Budget Headings	Total budget allocated	Expenditure this reporting period	Total expenditure to date	Further information
Staff	£91,931	£16,966	£25,771	This figure is lower than the £99,978 original budget because of variation in staff start dates. We have some underspend because the Software Developer post was vacant for longer than expected; we have additional costs for extra University of Leeds staff time plus additional costs as the end of the project has shifted.
Travel & Subsistence	£2,000	£839	£1,080	We have moved a little more funding into this category. Train fares have increased significantly and we have had an additional person linked to the project from Leeds (Systems Team).
Equipment	£4,000	£2,180	£3,500	
Dissemination activities	£1,300		£0	
Evaluation activities	-	-	-	
Other: software upgrade and associated consultancy	£2,820	-	£2,820	
Directly Allocated	£81,000	20,833	£40,500	
<b>Total</b>	<b>£183,051</b>	<b>£40,818</b>	<b>£73,671</b>	
JISC	£91,489			
Partners	£91,562			

## Appendix B: Revised Milestone Dates

### IncReASe Milestones

Milestone	Date	Deliverable	Staff
1	July 07	project plan, ( including work-packages plan, evaluation plan, quality assurance plan, dissemination plan, exit/sustainability plan)	PM
1	July 07	JISC site web page & Project web site	PM, TO
2	31/10/07	Two project reports per year (Oct 07)	PM
3	30/11/08	Database import case studies	PM,TO
4	15/12/08	Pilot test report.	PM, PO, TO
5	31/05/08	Web scraper perl script	TO
6	28/02/08	Review of scholarly works application profile implementation	PO, PM
7	30/04/08	Two project reports per year (April 08)	PM
8	31/10/08	Self-archiver case studies	PM, PO
9	30/11/08	Metadata enhancement report	PO, PM, TO
10	25/04/08	Database prevalence report.	PO
10b - new	31/03/08	Online questionnaire for researchers	PO, PM
10c - new	31/05/08	Questionnaire analysis	PO,PM
11	30/06/08	ESRC workflow report	PM, PO, TO
12	31/01/09	Report on workpackage2, evaluating workflows.	PM
13	31/08/08	Research funder workflow scenarios.	PM, PO
14	31/08/08	Services interviews report	PM, PO
15	31/01/09	Guidelines for bulk import	TO, PM
16	31/10/08	Two project reports per year (Oct 08)	PM
17	31/12/08	Centralising deposit report.	PM, PO
18	30/11/08	Workflow for local systems: roles and scenarios	PM
19	31/01/09	Impact Report	PM, PO
20	15/02/09	WP 3 Technical Report	TO, PM
21	28/02/09	Automation issues report.	TO, PM
22	28/02/09	Web site status clear. Available for at least 3 years.	TO, PM
23	28/02/09	Completion Report	PM, PO,TO
24	28/02/09	Final Report	PM, PO,TO

Black = complete

Grey = partially complete

Red= milestone date moved to later in the project

## **Appendix C: Requested changes to the JISC web site**

New project end date: 28/02/09

### **Project Manager**

The correct telephone number is 0113 343 7067

### **Project Team**

Beccy Shipman, Project Officer

0113 343 7067

Archana Sharma-Oates, Software Developer

0113 343 7067