

Project Identifier: UKRDDS
Version: 1.0
Contact: Michael Whitton
Date: November 2016



Jisc Final Report

Project Information			
Project Identifier	UKRDDS		
Project Title	UK Research Data Discovery Service		
Project Hashtag	#JiscRDDS		
Start Date	30 March 2015	End Date	30 Sept 2016 (revised from 31 July 2016)
HEI/Data Centre	University of Southampton		
Main Contact	Michael Whitton		
Contact email	M.Whitton@soton.ac.uk		
Jisc Project Manager	Christopher Brown		

Document Information			
Author(s)	Michael Whitton, Adrian Cox		
Project Role(s)	Project Lead		
Date	30/11/16	Filename	UKRDDScfinalreport_Southampton.v1.0
Access	This report is for general dissemination		

Document History		
Version	Date	Comments
1.0	30/11/2016	

Table of Contents



- 1** **ACKNOWLEDGEMENTS**
- 2** **PROJECT SUMMARY**
- 3** **MAIN BODY OF REPORT**
 - 3.1** **PROJECT TEAM**
 - 3.2** **REQUIREMENTS**
 - 3.3** **PROJECT OUTPUTS AND OUTCOMES**
 - 3.4** **ISSUES AND CHALLENGES**
 - 3.5** **WHAT DID YOU LEARN?**
 - 3.6** **IMMEDIATE IMPACT**
 - 3.7** **FUTURE IMPACT**
- 4** **CONCLUSIONS**
- 5** **RECOMMENDATIONS**
- 6** **IMPLICATIONS FOR THE FUTURE**
- 7** **REFERENCES**

1 Acknowledgements

We thank Jisc for funding the University of Southampton to participate in this UKRDDS project.

2 Project Summary

The University of Southampton engaged with the UK research data discovery project¹ in the first phase, helping to develop metadata profiles. In the second phase we continued to work with integration of our Institutional Repository, and developed a use-case to enable more effective discovery of repository endpoints.

Contributing metadata from ePrints Soton

We have contributed metadata for datasets from our Institutional Repository, ePrints Soton. We also engaged with Jisc to improve their metadata specifications, and the mapping of our metadata, which raised the weakness of lack of consistent implementation of the Dublin Core standard.

Looking forward we are implementing Pure as our Current Research Information System (CRIS) and retaining ePrints Soton as the discovery portal. This will maximise both reporting capabilities and discoverability, which will be further enhanced via the UKDDS. We will also investigate improving the metadata that is exposed to services like the UKRDDS.

We recommend joint working within the EPrints community to produce a plugin to enable any EPrints repository to expose metadata in DataCite 4.0 format. This will enable repositories to easily and effectively participate both with the UKRDDS and OpenAire.

Usability of the alpha UKRDDS

We also assessed the usability of the alpha version of the Research Data Discovery Service (UKRDDS)². This was both from the perspective of people discovering our data, and discovering data deposited in national/disciplinary repositories.

We recommend Jisc continue to improve both the usability – both in terms of the interface and exporting/harvesting options.

Use case for auto-discovery of repository endpoints

We have also developed a technical use case for auto-discovery of OAI-PMH Endpoints via an institution's Organisation Profile Document (OPD)³. We have researched how institutions currently add details of publication and dataset repositories, and how this could offer a sustainable way to discover and manage repository end points and contacts. We recommend that the OPD terms are rationalised and scope notes revised to improve their consistency and usefulness.

3 Main Body of Report

3.1 Project Team

Name	Email	Role
Michael Whitton*	M.Whitton@soton.ac.uk	Project lead (repository)
Adrian Cox*	A.J.Cox@soton.ac.uk	Project lead (equipment.data/OPD)

¹ Jisc, "Uk Research Data Discovery: Making Research Data Discoverable (R & D Project)," <https://www.jisc.ac.uk/rd/projects/uk-research-data-discovery>.

² Jisc, "Research Data Discovery Service (Alpha)," <http://ckan.data.alpha.jisc.ac.uk/dataset>.

³ data.ac.uk, "Organisation Profile Documents: What Is an Organisation Profile Document (OPD)?," <http://opd.data.ac.uk/>.

Andrew Milsted	A.Milsted@soton.ac.uk	Project officer (equipment.data/OPD)
Dorothy Byatt	D.R.Byatt@soton.ac.uk	Project officer (repository)
Isobel Stark	I.A.Stark@soton.ac.uk	Project officer (repository)
Wendy White	W.H.White@soton.ac.uk	Strategic oversight

The two project leads principally represented the University in Advisory Group meetings, with support from the rest of the team to contribute to as many of the meetings as possible.

3.2 Requirements

We have a single Institutional Repository containing both publications and datasets. This combined approach is more unusual, with advantages of making it easier for academics to interact with a single system. There are challenges with visualising more heterogeneous data. We are currently implementing Pure as our Current Research Information System (CRIS), which will enhance reporting capability for senior managers assessing a range of research outputs. The data will still be exposed via our EPrints repository – this is key for discovery, which will be enhanced by inclusion in the UKRDDS.

We required harvesting only the datasets (identified by a specific OAI-PMH set), the service was able to do this, and regularly add new items, update metadata and delete records as appropriate. The service was able to map most of the metadata we expose and visualise it. Two fields were missing – a link to the repository record (due to differences in how the EPrints repository software’s default implementation of Dublin Core; and DOI which was only exposed within the citation.

Our researchers do also deposit data with national data centres such as the UK Data Service, the Archaeology Data Service and NERC data centres. It is possible to search for datasets from our authors (by searching for “university of Southampton” in all fields); however there are difficulties in doing a more sophisticated search:

- This information is in different fields (depending on the source repository)
- It is difficult to search specific fields. The field name needs to be entered into a free text box, so it is easy to type the wrong name, and searches on creator do not work.
- Only limited functionality is provided to export the records (in JSON format).
 - This only returns a limited number of records.
 - Also the JSON export effectively does it’s own search ignoring any filters that have been applied (e.g. on Organisation/Repository).
 - Alternative formats like CSV/Excel would make this feature more widely usable to those without the skills to manipulate JSON format
- Also there is no facility to harvest records by OAI-PMH etc.

Delivering these improvements by Jisc with community engagement will add value to the service enabling institutions to better understand their institutions research data.

3.3 Project Outputs and Outcomes

Output / Outcome Type (e.g. report, publication, software, knowledge built)	Brief Description and URLs (where applicable)
Technical use case	Research outputs discovery using the Organisation Profile Document (OPD) http://doi.org/10.5258/SOTON/402662
Blog post	Measuring equipment utilisation – could impact be the answer? http://www.encyclicexchange.ac.uk/9445/measuring-equipment-utilisation-impact-answer/
Blog post	Post project reflections (forthcoming).

	https://unisouthamptonlibrary.wordpress.com
Knowledge built	Metadata harvesting – improved knowledge across the community of how to effectively harvest metadata from research data repositories.
Knowledge built	Expanded team knowledge of research data management including shared community learning across all the project participants.
Service development	Added value of this service embedded in University of Southampton researcher training to incentivise research data deposit.

3.4 Issues and Challenges

The initial challenge was to ensure only datasets were harvested (i.e. not publications) – this was easily achieved by supplying the ‘setspec’ of the appropriate OAI-PMH set. This identifier allows the harvester to target a collection of records with the item type ‘dataset’.

Another challenge was the metadata in our OAI-PMH feed which was originally designed for publications. Although the key metadata elements are present, a number of fields are missing. Also we exposed the weakness of the lack of consistency of interpretation of Dublin Core and subsequent impact on harvesting for data. This has also impacted on other EPrints repositories in the project.

Also a challenge came from the implementation of our CRIS, Pure. This will change the way metadata is collected, though ePrints Soton will still be the discovery portal. This will affect how metadata will be collected in the future. We have prioritised ensuring all key metadata will pass from Pure into ePrints Soton. Once Pure is live we can consider how metadata can best be visualised and harvested.

- The project helped show us the challenges of working with proprietary CRIS systems and metadata exchange for data, linking through to open systems for discovery - both institutional and national.
- Importance of working with CRIS suppliers as well as developing open systems for an integrated landscape and researcher workflows

Our technical use case for auto-discovery of OAI-PMH endpoints was developed in draft fairly early in the project. Feedback from Jisc, funders and/or other institutions would be very useful to help the development process of this use case.

3.5 What did you learn?

We engaged with the discussions over metadata standards, to help us overcome the issues identified in 3.4. This raised our awareness of different practices and needs across institutions, and the challenges of designing a flexible solution for them all. Initial discussions implied the preferred result in the longer term was a bespoke schema for the UKRDDS (with mapping to existing schemes as an interim solution), however this developed into flexible requirements based on existing standards. Thus a standard like DataCite could allow harvesting by the UKRDDS, and other international resources like OpenAire.

We found the initial stage of the project looking at use cases engaging, however it was easier to assess these once the Alpha Ckan repository was in place.

We were able to design a technical use case for auto-discovery of OAI-PMH endpoints. However encouraging institutions to add this information to their OPD has been less successful than expected. The equipment.data team were able to promote addition of OAI-PMH end points to the OPD as part of

their service engagement – leveraging participation in the UK RDDS project. This saw a noted interest in the benefits of this addition.

Institutional contact through the Extending the OPD to cover RDM⁴ project funded by Jisc has seen some engagement. We look forward to this being taken forward with the stakeholders for future development. Currently there are seven institutions listing a publications repository, and four with a data repository. There are two OPD terms for publication repositories and three for data repositories. There are variations and inconsistencies in how they are used.

Due to the way these questions are worded we had to refer to our Legal Services department, which has significantly delayed our engagement with this part of the OPD. This demonstrates that time needs to be factored in during planning for institutional sign off processes - even for something that may appear simple. Also there was some ambiguity in the questions, especially the three relating to data repositories. Rewording the questions would help to make the intended answer more clear.

Term	Description	No of Institutions
datasets_research_outputs	Research outputs dataset	8*
lyou_research_publications_repository	Institutional publications repository page	2
lyou_research_data_repository	Active data storage page	2
lyou_research_data_catalogue	Data registration and discovery mechanism page	3
lyou_research_data_repository_long_term	Longer term data repository page	1

This is based on Data from 4th November 2016.

* These include 7 publication/mixed repositories and 1 data repository.

A better strategy may be to encourage enrichment of metadata in ROAR⁵ and/or Re3data⁶ - after ensuring appropriate metadata fields are present.

3.6 Immediate Impact

The project has improved knowledge exchange between the library, and staff in Research & Innovation services working with equipment.data. The partnership also helped to develop the technical use case for auto-discovery of OAI-PMH endpoints. This has provided Jisc with a mechanism for the sustainable discovery and managing of repository end points and the responsible contact.

The project was helpful for research engagement and promoting importance of research data as part of strategic thinking and research practice.

This also helped us problem solve issues relating to our Pure implementation, informing forward planning for future developments and sharing learning with Research Office and IT colleagues. This has also assisted us in early lobbying of Elsevier in terms of product development and community user group engagement to improve functionality for data.

⁴ Digital Curation Centre, "Extending the OPD to Cover Rdm," <http://www.dcc.ac.uk/projects/opd-for-rdm>.

⁵ University of Southampton, "Registry of Open Access Repositories," <http://roar.eprints.org>.

⁶ re3data.org Project Consortium, "Registry of Research Data Repositories," <http://www.re3data.org/>.
UKRDDS Final Report Template v1.0

3.7 Future Impact

We intend to improve the exposure of metadata from our repository. This is part of a planned program of work to improve the Discoverability of our repository. We will consider migrating the dataset metadata we release through OAI-PMH from Dublin Core to DataCite 4.0 to enable us to make as much metadata available as possible. We will work with the EPrints software community where changes to plugins etc. may be needed to accommodate DataCite 4.0.

Once our CRIS, Pure is implemented we can explore equipment.data integration further. Academics will be able to link publications, datasets and grants to equipment. We will be able to monitor uptake of this facility through running reports in Pure.

This will help us model whole lifecycle modelling of metadata flows, including archival data with our Arkivum implementation as the volume and heterogeneity of data we engage with increases.

4 Conclusions

We have found the UKRDDS a useful way to allow people to search and visualise UK research data.

For this service to be effective key issues are:

- Standards to enable transfer of metadata from the repository. The Dublin Core Metadata Element Set is a useful vocabulary that many repositories support, however standards like DataCite support exposure of a richer metadata profile.
- A user friendly interface enabling the service to be easily and effectively searched.
- Enabling re-use of the data by providing options to download and harvest the metadata.

We have developed a technical use case for auto-discovery of OAI-PMH endpoints. This is a very viable process for sustainable end point discovery however, more work is needed to develop a source with widespread institutional engagement, possibly requiring further project engagement and/or policy/guidance leverage e.g. through RDM guidance.

5 Recommendations

We recommend an EPrints plugin to expose metadata in DataCite 4.0 format is developed by the EPrints community, with support from other interested bodies including Jisc and the British Library. This will enable any data repository using the EPrints software to more effectively contribute metadata to UKRDDS and OpenAire. This will also reduce duplication of effort by individual institutions.

We recommend Jisc continue to improve the UKRDDS with engagement and support from the community – both the quality of the interface and export/harvesting options.

We recommend Jisc investigate possible harvesting of relevant preservation metadata for discovery and link to potential national or endorsed network of preservation and archiving services.

Autodiscovery of OAI-PMH end points via the OPD would provide a robust sustainable resource for the aggregation and management of end point locations and the institutional responsible contact. Further consideration should be given as to the benefits of this approach and mechanisms for encouraging publishing via the OPD.

We recommend the OPD terms for publication and data repositories are rationalised and improved scope notes written to improve consistency of their use.

6 Implications for the future

The work on auto-discovery of endpoints could be continued by:

- Encouraging more institutional engagement with the Research Data Management parts of the 'Linking You' section of the OPD.
- Work on enriching metadata in ROAR and Re3data, to develop this as an alternative source of data.
- Funder policy/guidance for improving access to research outputs. This should be explored in discussion with RCUK

As part of the wider developments from the transition to a full service and sustainability models, there needs to be consideration how this development fits into the wider research infrastructure vision from Jisc. We would welcome discussion exploring the balance between community engagement and input and core central Jisc services.

The University of Southampton is happy to continue our participation with the UKRDDS, and to have metadata from ePrints Soton harvested.

The long term contact will be: Michael Whitton.

7 References

Cox, Adrian. "Measuring Equipment Utilisation – Could Impact Be the Answer?" *Efficiency Exchange*. 6 June, 2016, <http://www.encyexchange.ac.uk/9445/measuring-equipment-utilisation-impact-answer/>.

———. "Research Outputs Discovery Using the Organisation Profile Document (OPD)." University of Southampton, 2016. <http://doi.org/10.5258/SOTON/402662>.

data.ac.uk. "Organisation Profile Documents: What Is an Organisation Profile Document (OPD)?" <http://opd.data.ac.uk/>.

Digital Curation Centre. "Extending the OPD to Cover Rdm." <http://www.dcc.ac.uk/projects/opd-for-rdm>.

Jisc. "Research Data Discovery Service (Alpha)." <http://ckan.data.alpha.jisc.ac.uk/dataset>.

———. "Uk Research Data Discovery: Making Research Data Discoverable (R & D Project)." <https://www.jisc.ac.uk/rd/projects/uk-research-data-discovery>.

re3data.org Project Consortium. "Registry of Research Data Repositories." <http://www.re3data.org/>.

University of Southampton. "Registry of Open Access Repositories." <http://roar.eprints.org>.