



## JISC Final Report

Project Information			
<b>Project Identifier</b>	<i>To be completed by JISC</i>		
<b>Project Title</b>	DataPool		
<b>Project Hashtag</b>	#Jiscdatapool		
<b>Start Date</b>	1 October 2011	<b>End Date</b>	31 March 2013
<b>Lead Institution</b>	University of Southampton		
<b>Project Director</b>	Wendy White		
<b>Project Manager</b>	Dorothy Byatt, Steve Hitchcock		
<b>Contact email</b>	whw@soton.ac.uk		
<b>Partner Institutions</b>			
<b>Project Web URL</b>	<a href="http://datapool.soton.ac.uk/">http://datapool.soton.ac.uk/</a>		
<b>Programme Name</b>	Managing Research Data Programme 2011-13		
<b>Programme Manager</b>	Simon Hodson		

Document Information			
<b>Author(s)</b>	Wendy White, Steve Hitchcock, Dorothy Byatt		
<b>Project Role(s)</b>	Project Director, Joint Project Managers		
<b>Date</b>	12 April 2013	<b>Filename</b>	datapool-finalreportv03.doc
<b>URL</b>	<a href="http://eprints.soton.ac.uk/352949">http://eprints.soton.ac.uk/352949</a>		
<b>Access</b>	This report is for general dissemination		

Document History		
Version	Date	Comments
0.1	6 Feb 2013	First draft with headline considerations and outputs for each section - for ongoing team review (SH)
0.2	4 March 2013	Second draft for ongoing team review (additions by WHW/DB)
0.3	30 April 2013	Third full written draft (WHW)
4	14 May 2013	Final version incorporating feedback from project members (WHW)

## Table of Contents

NB : This table of contents 'auto-populates' - to update the table of contents – place cursor in the table of contents, right-click your mouse, click 'update field', select appropriate option

<b>1</b>	<b>ACKNOWLEDGEMENTS .....</b>	<b>3</b>
<b>2</b>	<b>PROJECT SUMMARY .....</b>	<b>3</b>
<b>3</b>	<b>MAIN BODY OF REPORT .....</b>	<b>4</b>
3.1	PROJECT OUTPUTS AND OUTCOMES .....	4
3.2	HOW DID YOU GO ABOUT ACHIEVING YOUR OUTPUTS / OUTCOMES? .....	5
3.3	WHAT DID YOU LEARN? .....	8
3.4	IMMEDIATE IMPACT .....	10
3.5	FUTURE IMPACT .....	10
<b>4</b>	<b>CONCLUSIONS .....</b>	<b>11</b>
<b>5</b>	<b>RECOMMENDATIONS .....</b>	<b>12</b>
<b>6</b>	<b>IMPLICATIONS FOR THE FUTURE .....</b>	<b>12</b>
<b>7</b>	<b>REFERENCES .....</b>	<b>13</b>
<b>8</b>	<b>APPENDICES (OPTIONAL) .....</b>	<b>ERROR! BOOKMARK NOT DEFINED.</b>

## Acknowledgements

Funded by Jisc within its Managing Research Data Programme 2011-13, DataPool is an institutional project for the University of Southampton, and has benefitted from the work and advice of many participants across the institution contributing a range of knowledge and skill, including disciplinary data experts and Faculty contacts. We particularly want to acknowledge the major contributions from those listed below, including our collaborators from outside the University and on the Steering Group.

The project investigators: Wendy White (Head of Scholarly Communication, Library), Mark Brown (University Librarian), Les Carr (Director, Web Science Doctoral Training Centre), Simon Cox (Associate Dean Enterprise, Faculty of Engineering and the Environment and Head of Computational Engineering and Design), Graeme Earl (Chair of Digital Economy University Strategic Research Group), Peter Hancock (Director of iSolutions), Jeremy Frey (Professor of Physical Chemistry and Head of Structure and Materials)

Project managers: Dorothy Byatt (Co-Project Manager, Library), Steve Hitchcock (Co-Project Manager, Electronics and Computer Science)

Expert contributors: Gareth Beale and Hembo Pagi (Archaeology), for imaging case study and repository testing; Mark Scott (Engineering Sciences/iSolutions), Gareth Beale (Archaeology) and Sandy MacKinnon (iSolutions) for developing training; Simon Coles, Philip Adler, Andrew Milstead (Chemistry) for testing and developing DataCite use; Arkadiusz Wisinowski (Centre for Population Change), Martin Chivers (iSolutions) and Michael Whetton (Library) for data visualization and impact case study; Peter Gibbs, Rich Gooding and Chris Yorke (iSolutions) for SharePoint Development; Patrick McSweeney, Matt Smith (iSolutions) and Tim Brody (EPrints Services) for EPrints development; Research Data @Essex team, including Louise Corti, Tom Ensom and Alexis Wolton for EPrints app; Oz Parchment, (iSolutions) for work on data storage requirements, Meriel Patrick (Oxford) for collaboration on training survey.

Steering Group: Louise Corti (Associate Director, UK Data Archive), Graham Pryor (Associate Director, Digital Curation Centre), Sally Rumsey (Digital Research Manager at The Bodleian Libraries, University of Oxford), Helen Snaith (National Oceanography Centre, Southampton), Philip Nelson (Pro-Vice Chancellor Research), Adam Wheeler (Provost and Deputy Vice Chancellor), Mylene Ployart (Associate Director, Research and Innovation Services) and the project investigators.

All Associate Deans for Research, Faculty contacts and members of the Researcher Development Graduate Centre who helped iterate policy/guidance and promote developing services.

Simon Hodson, Jisc Managing Research Data Programme Manager and Laura Molloy, Humanities Advanced Technology and Information Institute, University of Glasgow for their expertise and support throughout the project.

## 1 Project Summary

Calls for increased access to the results of research, both publications and now the supporting data, for the purposes of wider reuse and validation, have led to emerging new policy directions from research funders and institutions. The University of Southampton is developing its approach to research data management services, and in 2011 established a 10-year roadmap to guide this process, through the JISC-funded Institutional Data Management Blueprint (IDMB).<sup>1</sup>

The aim of the DataPool project was to deliver the first phase of that roadmap, in seeking to build capacity within the University to support effective data management practice across all disciplines, including multi-disciplinary activity, throughout the data lifecycle.

This is achieved by:

- clear institutional policy for research data management
- new embedded support services and guidance for researchers, beginning with bid preparation and submission
- tested training materials and workshop delivery model for postgraduates and early career researchers
- enhanced repository infrastructure to create comprehensive records of data outputs, provide locally managed storage of small-scale outputs and a platform for sharing data
- Case studies to investigate identified multidisciplinary issues in depth

These elements reflect the core components of the first phase (years 1-3) of the Roadmap and provides a foundation for looking ahead to the medium term.

Critically, through the project the University has established and will maintain a network of senior managers, disciplinary leaders, faculty contacts and data producers, both through groups established to support DataPool and existing networks such as the multidisciplinary University Strategic Research Groups (USRGs), to both inform and embed agreed data practices across the university.

The project used the evidence base from IDMB to inform policy and service development, and provide a basis for the promotion of new embedded support services from data creation through to data publication and reuse. The Data Management Planning Service ResearchData@Soton is a good example of a service with growing uptake, which is providing significant additional evidence through the service logs to inform the next phase of development.

The DataPool project was as much about cultural as technical change, embedding previous pilots into long-term institutional services across disciplines and working on iterative development through a partnership approach.

## 2 Main Body of Report

### 2.1 Project Outputs and Outcomes

Output / Outcome Type (e.g. report, publication, software, knowledge built)	Brief Description and URLs (where applicable)
Research Data Management Policy	Adopted February 2012 and published in the official University Calendar <a href="http://www.calendar.soton.ac.uk/sectionIV/research-data-management.html">http://www.calendar.soton.ac.uk/sectionIV/research-data-management.html</a>
Policy implementation and engagement report	<a href="http://eprints.soton.ac.uk/351945">http://eprints.soton.ac.uk/351945</a>
Research data management guidance, web site and collected briefing documents	<a href="http://www.southampton.ac.uk/library/research/researchdata/">http://www.southampton.ac.uk/library/research/researchdata/</a>
Introducing research data, student guide and training document, with case study exemplars	<a href="http://eprints.soton.ac.uk/338816/">http://eprints.soton.ac.uk/338816/</a>
Professional staff support for research data management, training report and survey results	<a href="http://eprints.soton.ac.uk/352107/">http://eprints.soton.ac.uk/352107/</a> Comparative data with Oxford will be available later
Training model and workshops for Early Career and PhD Researchers: qualitative report and	<a href="http://eprints.soton.ac.uk/351026/">http://eprints.soton.ac.uk/351026/</a>

feedback	
Report and feedback from the Data Management Planning consultation service	<a href="http://eprints.soton.ac.uk/351027/">http://eprints.soton.ac.uk/351027/</a>
Training matrix map to RDF Info Lit lens/Brewerton Research Lifecycle	<a href="http://datapool.soton.ac.uk/2013/01/21/mapping-training-needs-for-the-support-team/">http://datapool.soton.ac.uk/2013/01/21/mapping-training-needs-for-the-support-team/</a> blog post with matrix file
Multidisciplinary Imaging Case study: raster and 3D data	<a href="http://eprints.soton.ac.uk/350738/">http://eprints.soton.ac.uk/350738/</a>
Case study: Data visualisation and impact, ESRC Centre for Population Change	Integrated Modelling of European Migration Database <a href="http://eprints.soton.ac.uk/350672/">http://eprints.soton.ac.uk/350672/</a>
Case study: Tweet archiving	<a href="http://eprints.soton.ac.uk/350646/">http://eprints.soton.ac.uk/350646/</a>
Cost-benefits: analysis of case studies	<a href="http://datapool.soton.ac.uk/2013/03/21/cost-benefit-analysis-experience-of-southampton-research-data-producers/">http://datapool.soton.ac.uk/2013/03/21/cost-benefit-analysis-experience-of-southampton-research-data-producers/</a>
EPrints data app implemented in Soton Service	Live service on 11 April 2013 <a href="http://eprints.soton.ac.uk/">http://eprints.soton.ac.uk/</a> using ReCollect app <a href="http://bazaar.eprints.org/">http://bazaar.eprints.org/</a>  Champion depositors identified for early engagement - June onwards
Repository data deposit workflows – report	<a href="http://eprints.soton.ac.uk/352813/">http://eprints.soton.ac.uk/352813/</a>
Sharepoint test data deposit workflows	Resource available internally via Sharepoint at Southampton as working demonstrator, full implementation dependant on overall institutional timescales for SharePoint services. Use as document management service now consolidated, but this was later than planned.
Link Arkivum A-Stor archive to ePrints	Integration of Arkivum's A-Stor archive service with ePrints - Bazaar app <a href="http://bazaar.eprints.org/">http://bazaar.eprints.org/</a> with user documentation
Proof of Concept demonstrator of test cases	Working use cases for file replication and transfer.
DataCite DOI applications	DataCite DOI functionality embedded in eCrystals data repository App for automating DOI minting in EPrints repositories <a href="http://bazaar.eprints.org/">http://bazaar.eprints.org/</a>  Demonstrator DOI assignment for LabTrove notebooks

These outputs address training and advocacy and demonstrate progress with the evaluation and implementation of some key technical elements. In-depth investigation and exemplars in disciplines which are of particular significance for the University provide impact narratives and business analysis to help long-term planning.

## ***2.2 How did you go about achieving your outputs / outcomes?***

The aim of the DataPool project was to build on the researcher-led Institutional Data Management Blueprint Project (IDMB). This provided an evidence-base through surveys, interviews and use of AIDA<sup>ii</sup> to develop an aspirational 10 year institutional roadmap with associated cost modelling and some initial testing of ideas and tools through disciplinary activity. The purpose of the DataPool project was to deliver the first phase of the roadmap and realise the aspiration through the

development of policy and underpinning services at institutional scale. The ambition was to move from individual initiatives to embedded support, with a focus on cultural change and partnership-building to improve research data management practice. This would be achieved by embracing existing informal networks such as the multi-disciplinary University Strategic Research Groups (USRGs) and utilising existing communication routes between professional services and the academic community to develop robust services, recognising the complexity of the stakeholder environment for the life-cycle of research data management emphasised by the findings of the IDMB surveys.

Southampton has long standing experience of interfacing with sector-wide service providers and collaborations, so it was also important for the project to build on linked activity including leadership of the EPSRC IT as a Utility Network, Directorship of the UK National Crystallography Service, established expertise in High Performance Computing and e-Science, initiatives in open and linked data and work with National Data Centres. This element of project engagement was facilitated through the involvement of senior academic Co-Investigators in these areas and the extremely useful advice from project Steering Group representation from the British Oceanographic Data Centre and the UK Data Archive.

There were four main themes of activity:

- Implementation of an institutional research data management policy with an associated one-stop-shop of web guidance and data management planning advisory service;
- Development and delivery of training for research students and staff with associated case studies which integrate into a range of learning contexts;
- Development, testing and implementation of repository and storage options leading to the enhancement of data catalogue services;
- Investigation of more specialist multi-disciplinary academic challenges to provide depth of understanding and inform implementation of service enhancements, on-going cost analysis and business decisions – there was particular focus on raster and 3D imaging requirements and data visualisation and impact.

These strands were all pursued in parallel and iteratively, recognising their interdependence and maximising the benefits of cross-fertilisation.

The approach of the project was to build on engagement with formal and informal fora, continuing to work with senior management with both the Provost/Deputy Vice-Chancellor and the Pro-Vice Chancellor for Research on the Project Steering Group and working with a broader range of researchers across many more disciplines to develop and embed services. There was clear early steer from all quarters that this was a project that was as much about enhancing research practice and research impact than about compliance, although the importance of the evolving policies of research funders was recognised. In particular the EPSRC requirements for data storage and retention with provision of a data catalogue and strategic roadmap clearly dovetail with the aims of the project.

Engagement with a wide range of researchers was essential to unravelling the complexity of the issues and allowed for nuanced disciplinary and even-sub-disciplinary perspectives. This was particularly evident as drafts of the Research Data Management Policy were discussed. Nominated Faculty champions liaised with colleagues and raised issues relating to appraising the significance of data: ethics of data as objects and impact on the scope of what data come under the terms of the policy, the relationship between requirement and service provision, the assignation of roles and responsibilities for decisions across the data life-cycle and in the longer term. The policy was passed by Senate in February 2012 and the associated one-stop-shop web guidance and Data Management Planning Service was launched. The policy is not seen as set in stone and the expectation from senior staff, researchers and services is that this is an iterative process with guidance maturing in response to feedback and relevant amendments are likely to be made to the policy after reflection on early implementation. There is recognition that services are still evolving to support the policy and that in many ways the policy provides a helpful framework for development as well as providing a commitment to principles.

All aspects of the project have involved partnership working, not just to deliver individual project outputs but to set up long term services. This involved identifying relevant expertise in existing roles,

identifying gaps in expertise or in critical mass of support and evolving knowledge in emerging roles. It also required use of a variety of communication methods across the full range of services. Service and academic partnerships were particularly crucial for the development of the web-based guidance and the data management planning (DMP) support service with e-mail, face-to-face and Skype/telephone consultation options. The Library leads the triaging of the advertised service, however in reality because of the many perspectives and priorities for research data management, queries can start out at any service point. Raising awareness of the service with many groups of staff has been important to harness the most effective communication routes and ensure that the best advice is given.

Key roles are the Faculty Business Relationship Managers for the central IT services (iSolutions), Academic Liaison staff in the library supporting specific disciplines, Collaboration Managers, Bid Managers and Research Support Officers in Research and Innovation Services who due to their involvement with the bid process can be early referrers but also contributors in specific areas such as intellectual property. A log is being kept of consultations to provide an evidence base for training staff and for the development of anonymised exemplars. This has confirmed our expectation that in most cases more than one service perspective is required to work through issues and in some cases academic experts are also able to assist with specific issues. The DMP service has been well received and led to direct requests for specific DMP training for all staff (from early career researchers to PIs) in three Faculties – Medicine, Health Sciences and Natural and Environmental Sciences. This training has been co-delivered by iSolutions and the Library and illustrates the interlocking nature of the strands of the project with policy and service development informing training and training feedback informing services. There is a developing programme of training for professional services staff which will be further influenced by the results of a survey exploring levels of staff confidence in various research data management areas and knowledge of referral options. The survey has also been completed by staff in Oxford which should provide useful comparative data and opportunities for joint/regional training.

Co-delivery of training has involved another significant contributor to the project, PhD researchers and technical experts working in Faculties. The approach to training has focussed on supporting PhD researchers and as such peer involvement has been important for engagement. The project has delivered generic “first principles” training through the Researcher Development Graduate Centre (RDGC) advertised courses and more focused workshop style training by specific Faculty and through a Doctoral Training Centre (DTC). PhD researchers have been involved in developing the supporting guidance, disciplinary case studies and in co-delivering training. There has also been a pilot making data available to examiners during the PhD internal upgrade process which had positive feedback. The formal process for deposit of e-theses, which is mandatory at Southampton, is being amended to include the optional deposit of data in ePrints alongside the thesis. This has the full support of the RDGC and has been demonstrated and discussed in the Web Science DTC training event to obtain early feedback.

PhD researchers, research fellows and technical experts have been leading strands of the project that aim to promote an “innovation to service” approach that ties in with the mid-phase of the institutional roadmap. The multi-disciplinary case studies on imaging requirements and data visualisation for impact both provided evidence of innovation that can be embedded in services. The technical work which has provided EPrints bazaar apps to automate DataCite DOI minting and link EPrints to Arkivum storage gives us an opportunity for community implementation and service refinement. Further proof of concept work with DOI minting for Labtrove electronic notebooks and data transfer to Arkivum pave the way for possible shared services approaches with roles for third party providers.

Working with existing institutional and researcher workflows was seen as important. The rationale for the development of repository services was to work with two existing platforms with expertise and traction with the academic community and to continue to develop to potential of both to support research data management. We worked to a potential complementary model where SharePoint could offer a facility to manage data during a project and EPrints would have enhanced functionality to support datasets, particularly open datasets and data supporting academic publication.

The scale of the JISC RDM Programme provided many opportunities to work collaboratively and it was part of the approach of the project to maximise these opportunities. As well as specific examples

of collaboration outlined in this report, other significant areas of interest were the excellent training outputs of other programmes which informed our approach and exploring data citation issues. We particularly incorporated ideas from DataTrain<sup>iii</sup>, DaMSSI<sup>iv</sup> and RDMRose<sup>v</sup> for training which enabled us to map skills and attributes to the Researcher Development Framework. This enables links to other graduate programmes and helps identify and investigate areas for more support. We have used the material so far from PREPARDE<sup>vi</sup> for data citation as we develop our DOI service and links to publication models.

Engagement with other sources of support and sector opportunities has also been important e.g. use of the DCC tools like DMP Online<sup>vii</sup> which we have been using as a mediated support tool by services staff to support consultation with researchers. The British Library workshops on DataCite<sup>viii</sup> were highly valuable, informing the launch of our DOI service.

## **2.3 What did you learn?**

### **Research Workflows**

Where we were engaging with specific service improvements with a ready user base and at relatively low-cost, it was easier to follow the desired “innovation to service” model. For example the automated tools to support minting of DataCite DOIs have been developed by Chemistry and will be implemented first using the Crystallography repository where they already have expertise of using CrossRef DOIs and have publications which they need to relate to underlying data. Where we were working with the long-term roll out of a significant new institutional system e.g. SharePoint the timescales and development were more challenging. The aim was for SharePoint to offer a facility to deposit, manage and share data with colleagues as appropriate during the life-cycle of a project. This would complement ePrints as a repository which makes data available in support of papers and other forms of research impact. The project has delivered a pilot version of this SharePoint facility which was user tested by a group of PhD students and PIs across a wide range of disciplines. However due to delays and re-evaluation of the milestones for SharePoint roll out, implementation has been put on hold. SharePoint has now been launched across the University for document management, but the business-case for the next phase is still being assessed. There is a tension between embedding research data management requirements into a large-scale institutional strategic IT deployment and flexibility to respond quickly to user feedback. It is clear from the Jisc Programme as a whole that seamless technical workflow through the stages of the research lifecycle across all disciplines and at scale is a sector-wide challenge.

There is a widely recognised difficulty in providing an easy to use institution-wide repository and data catalogue and the desire to cater for disciplinary specificity, in particular additional metadata that facilitates re-use. Work with the IDMB and DataPool projects has developed approaches that have been based on the researcher perspective gained through the interviews, case studies and policy engagements. The three level metadata model developed by IDMB has been further explored by Essex<sup>ix</sup> as part of this Jisc Programme and informed their development of the ReCollect EPrints data app. Discussions between Essex, Southampton, Glasgow, Leeds and EPrints Services looked at standards and fields mapping. There was agreement that local implementations should keep core metadata such as that required for DataCite and INSPIRE in common and a commitment to engage as an EPrints community on future developments in this area. The ReCollect app is now integrated into the live service at Southampton and we anticipate that use of this service in earnest will provide additional feedback to inform service enhancements.

DataPool has explored issues of identifiers with the DataCite trial. The work undertaken by Chemistry in relation to the LabTrove notebooks has investigated potential granularity of DOI links and how the landing page might work with the dynamic notebook environment, yet provide a snapshot to support a publication. We already have a notebook example supporting a publication with the launch of our live DataCite service. The experience of the crystallographers testing the DOI implementation has been invaluable, but the assessment at scale means that a likely outcome is minting DataCite DOIs for the whole Crystallography Service as the transition is made from use of CrossRef. There are no easy solutions here, two identifiers are not ideal but as DataCite is specifically set up for data and has



growing researcher-led engagement and sector support through the British Library, there is a commitment to move to the DataCite Service.

#### Research Support

Institutional services also need to address the relationship with external disciplinary services and the approach of DataPool with the enhancement of the institutional repository and with the development of policy and guidance has been to work as a complement to existing services, particularly those linked to funders. Institutions and funders both want to facilitate improved research data management practice and may offer services that support this, so we are working towards the same goal. This has been a particular issue for our new Data Management Planning Service (DMP) and the web-based guidance where the approach is to help interpret the funders' requirements and map issues to both funder and institutional services, referring on to additional specialists as needed. It is clear from the early feedback from researchers using this service that this is highly valued.

We found that the DMP service was also important for engaging more experienced researchers who are time pressured and can be harder to reach. Broad awareness of any policy is often an issue, so the reflective approach required to develop the DMP is itself a tool for engagement. From this awareness of the research data management policy, service offers for storage and visualisation and many areas of RDM support are surfaced in a way that is integral to research activity. This has already in the Faculties of Medicine, Health Sciences and Natural and Environmental Sciences led to requests for specific training for PIs on research data management plans, which in turn has provided examples for the training DataPool has been developing for PhD researchers. Researchers have sometimes started out looking for a template solution, so the project has anonymised some examples, but the feedback for the face-to-face or conversation based support has been very positive. Based on this evidence the project team feel strongly that this adds significant value and the resourcing for the consultation model is built into service planning as part of long-term embedding and potentially for business-cases for more specialist input.

#### Research Training

PhD students and Research Fellows have been significant contributors to this project, leveraging knowledge from specialist data support roles in disciplines or developing new roles. The peer engagement has meant that they have acted as key change agents, often bridging research groups through engagement with multi-disciplinary fora. They have led project research activity through the case studies, contributed to testing the SharePoint and EPrints developments, designed training material and led workshops often in conjunction with colleagues from services. They have also contributed specific technical expertise as part of service development. For PhD and Research Fellows the informal networks have played a key role in developing and engaging with communities of practice. This model relies on a balance of central service support and local engagement and the appropriate blend of both to maximise uptake by researchers has been important for the developments of all the services.

#### Research Investigation and Exemplars

The case studies have revealed some relatively "quick wins" to promote change in practice or reveal service gaps, for example the addition of raster and 3D equipment to the EPSRC equipment list to promote usage and return on investment. They have also provided specific examples of storage issues, providing a level of detail that is useful for business planning and a story telling narrative giving the context required to engage managers. Also of value to business modelling was the development through the project of some proof of concept exemplars with possible shared and third party services, which could lead to full service offers. This includes the development of an app linking EPrints to Arkivum's A-Stor archiving service tested against some use cases and the link through DataCite implementation to an emerging service model for LabTrove electronic lab notebooks. The JISC programme has inspired much discussion of possible shared-services solutions and this is an area where we seek to continue to explore options.

#### Evidence

Evidence gathering has been an important focus of this Jisc programme, with quantitative and qualitative information important to support long-term service development and illustrate return on investment. Our focus has not just been on project metrics but information that will support the maturing services through the 10 year span of the institutional roadmap and beyond. Feedback from

training, enquiry and consultation support has already been embedded into institutional quality assurance processes. Some elements will take time to develop e.g. datasets recorded in the data catalogue and made available through repositories, datasets issued with a DataCite DOI and citations of those data, datasets available that support publications. This is an area where sector-wide metrics will proliferate and we hope that the experience from the programme will help us continue to contribute to discussions about data metrics so that we are engaging with real impact and not counting figures for the sake of it, or worse out of context inviting misleading conclusions.

## **2.4 Immediate Impact**

The project has always has the continued backing of the senior staff in the institution, but the steer provided by the project in drawing together activity across disciplines and services has been noted and resulted in consolidation of governance and data management roles. To ensure that the next phase of our 10 year roadmap fully embeds the tenets of the policy into practice right across the institution, we are proposing at the next Research and Enterprise Advisory Group (REAG) that the Steering Group for DataPool evolves with a link to REAG to provide continuity of governance for research data management issues. The PI of the project based in the Library will continue to lead co-ordination of research data management services and initiatives as this has also been formalised into a longer term role, recognising the strategic importance to the institution.

During the project the Library has reconfigured the balance of its thematic activity within its matrix structure. There is now a dedicated group of 6 staff to take a lead on research data management support and many of the members of this group combine this thematic responsibility with disciplinary academic liaison activity, including the DataPool Co-Project Manager. This is roughly equivalent to 3FTE focussed on RDM services and the aim is to build a nucleus of expertise whilst improving baseline skills across all academic-focussed library roles.

The more agile software development through Eprints has proved the quicker win for implementation and immediacy of impact. The integrated structural embedding required by SharePoint which is cross-cutting research, teaching and administrative infrastructure is a more significant challenge. However the feedback across the Jisc Programme has been that the linked infrastructure to support all curation needs across all stages of the research lifecycle is a challenge.

Feedback from the training and data management planning consultation services is now embedded in the regular institutional quality assurance processes which will provide a growing evidence base and demonstrates assimilation into the institutional culture. The unsolicited demand for additional Data Management Planning disciplinary group workshops as a result of individual consultations, demonstrates an encouraging level of engagement by PIs, which provides a good foundation for further progress.

Recognizing the success of the training model, the next phase of business planning will explore with the Researcher Development Graduate Centre the potential for funding PhD students, though stipends or scholarships, research data management champions who would have a role developing and delivering training in line with the service/peer co-delivery model used in the project. This builds on successful national initiatives, such as the Mayflower Scholarships and institutional initiatives, such as the use of "digital champions" linked to the Digital Economy University Strategic Research Group and the Centre for Innovation in Technologies and Education.

## **2.5 Future Impact**

The proof of concept demonstrator linking EPrints with ArXivum, tested against use cases where files are moved or replicated and mapped against institutional retention review triggers may be of interest to the community as a possible candidate for a consortial approach and will definitely lead to business modeling to contribute to sector discussion.

We will track the recent implementation of the enhanced EPrints data deposit through monitoring the number of deposits, downloads of open data and data citation. This service was planned as an end of project launch, but we hope that known enthusiasts for early usage will help build content to assist with showcasing the facility. A number of Data Management Plans have already included use of ePrints as part of their approach. We will continue to share our experience with the EPrints community as we work together to further improve the service, ensure inter-operability with a growing range of target repositories including disciplinary repositories, national repositories such as the Research Outcomes System and ResearchFish and exchange with publishers.

The case studies explored open data issues including enhanced data visualization and engaging with target audiences; adding value to administrative data to support effective use of resources and promote multidisciplinary research collaboration. This evidence from the project will help inform to the long term institutional strategy to provide sector leadership in open and linked data, an area where there are acknowledged benefits but also socio-technical complexities.

As we have tested the DataCite DOI minting with electronic lab notebooks and explored the levels of granularity of citation, the dynamic data environment and options for content negotiation, there is the potential for future service development. The project has provided a DOI app for EPrints and exemplars of work with LabTrove including a forthcoming publication citing notebook data. This work has highlighted the potential for further development in this area and we look forward to working with British Library and the EPSRC IT as a Utility Network to progress this.

Working with Oxford to roll out the survey on professional service staff skills and confidence in aspects of research data management has provided data which can now be used for benchmarking if we run the survey again in the future. We plan to exploit the potential for collaboration with a joint DCC engagement and look at regional sharing and training as well as individual activities bespoke to institutions.

### 3 Conclusions

There were four main strands to the DataPool project and over the course of the project we found a significant degree of commonality of approach across Jisc RDM projects, which has aided community sharing and helped identify what has worked well and which areas are more problematic.

The main conclusions from the 4 themes of the project were:

- Iterative development of policy, guidance and technical tools is important as a mechanism for engagement as well as ensuring that the outputs are fit for purpose. The project outputs should not be seen as static, but under a process of continual refinement
- Working with an “innovation to service” approach helps to translate the work of talented enthusiasts with considerable expertise into services which support knowledge transfer and collaborative practice
- Case studies are essential to explore issues in sufficient depth and provide granular evidence for investment and service improvements
- PhD researchers, Early Career Researchers and technical specialists are key change agents and the majority of successful project activity was achieved through their input
- A balance between provision of generic training, guidance and metadata and meeting discipline-specific needs is required. There is a spectrum of support and it is important to engage on all points of this spectrum.
- The whole spectrum of work in both the areas of technical innovation and cultural change is developing very fast and there is a significant need for a shared approach.
- Discipline particularities are increasingly significant in terms of research culture and practice, data content and format, and levels of technology required. This has a particular relevance to interdisciplinary research.

Particular comparator themes have been shared approached to evidence gathering to facilitate benchmarking and discussions about potential shared services and approaches to technical and standards development.

## 4 Recommendations

### General Recommendation

We hope that these recommendations can be considered by all relevant stakeholders including Jisc, DCC, institutions, funders and other sector bodies to help inform the discussions on sector priorities.

### Recommendations for Jisc

- The Jisc Managing Research Data Programme was extremely successful in supporting sharing across projects in a trusted environment and building on this approach through support for targeted events for specific sector-wide issues would be valuable
- Jisc should consider how to promote within the sector the significant work attached to case studies such as those in DataPool to create a pool of knowledge and expertise of issues at discipline or cross-interdisciplinary level to support institutions with their planning.
- Partnership exploration with institutions and funders would help improve complementarity of policy, guidance and Data Management Planning support for researchers
- There is scope for shared services with third party providers, but this requires sharing some aspects of business modelling, proof of concept developments and community knowledge transfer to be effective and focussed support for promising models would be helpful. This could take the form of brokering roles and expert knowledge input from organisations like Jisc and the DCC as well as funding.
- Training initiatives for professional staff, including IT, research office and library staff, are also developing and there is clear scope for some regional and national training to supplement the institutional activity. The DCC and Jisc have a role here to help institutions share good practice and give focussed support in areas of identified difficulty.

### Recommendations for the Wider Community

- Use of iterative approach to policy development is helpful to engage academic and service communities with change and avoid perception of policy as primarily concerned with enforcement rather than an enabler
- Further work on modelling costs of data appraisal in relation to storage would be useful. There are good examples of analysis for some disciplines and data centres, but the picture is currently rather fragmented
- Training initiatives for PhD and Early Career Researchers are bedding down in a number of institutions and it would be useful to unlock the detail of activity, sometimes available through relevant intranets, through working with VITAE, UKCGE and new doctoral training networks.
- Developing a comprehensive research data management infrastructure to support storage, management and curation across the lifecycle, including where relevant, long term archival storage, has proved a challenge across all projects. Progress has been incremental and the establishment and growth of relevant user communities to support specific tools and services will be important.

## 5 Implications for the future

As a project focussed on institution-wide embedding of research data management policy, services and infrastructure, sustainability of all core developments have been built into the budgets and planning for the relevant University services. There are some areas that would benefit from additional investment and/or community support and these have been highlighted in this report. These areas come into various categories:

- Project investigations revealed previously unknown gaps in demand for services. e.g. data visualisation support

- Where the aim was to explore the potential for a service e.g. work with Arkivum where the likely costs of a full service would require further business modelling including shared services options
- Where an approach to implementation has been tested and proved successful, providing evidence to assist with an institutional business case e.g. PhD training champions
- Case studies linked to service development where demand is not yet fully known and feedback from the service use will determine institutional commitment e.g. Tweepository
- Areas of sector-wide concern e.g. data retention appraisal, long term storage and tools to support the curation and collaborative use of project data. Whilst the project has helped developed some support e.g. enhanced data repository facility, it is recognised that this is part of a much wider on-going challenge.
- Agile, lightweight and often inexpensive technical change has had, or has future capacity to have, significant impact. EPrints apps and tools like ZendTo can simplify or add value to workflows which incentivise researchers.

Areas of work where with more support we could contribute to sector-leading initiatives are:

- services to support DataCite DOI minting
- ongoing work with Arkivum if there is sufficient sector interest and we are aware of a growing group of Jisc MRD funded institutions exploring this possibility
- sharing approaches to training and Data Management Planning with further exploration of the most effective balance between generic, bespoke and disciplinary/sub-disciplinary support.

Long-term contact: Wendy White, Head of Scholarly Communication, University Library  
whw@soton.ac.uk

## 6 References

- 
- <sup>i</sup> Brown, Mark, Parchment, Oz and White, Wendy (2011) *Institutional data management blueprint*, <http://eprints.soton.ac.uk/196241/>
- <sup>ii</sup> Assessing Institutional Digital Assets, [http://aida.da.ulcc.ac.uk/wiki/index.php/Main\\_Page](http://aida.da.ulcc.ac.uk/wiki/index.php/Main_Page)
- <sup>iii</sup> University of Cambridge, DataTrain <http://www.lib.cam.ac.uk/preservation/datatrain/>
- <sup>iv</sup> Digital Curation Centre, DaMSSI <http://www.dcc.ac.uk/training/damssi>
- <sup>v</sup> University of Sheffield, RDMRose <http://www.sheffield.ac.uk/is/research/projects/rdmrose>
- <sup>vi</sup> PREPARDE Project, <http://proj.badc.rl.ac.uk/preparde/blog/>
- <sup>vii</sup> Digital Curation Centre, DMPonline, <https://dmponline.dcc.ac.uk/>
- <sup>viii</sup> British Library, DataCite Workshops Archive, <http://www.bl.uk/aboutus/stratpolprog/digi/datasets/workshoparchive/archive.html>
- <sup>ix</sup> Tom Ensom, Adapting EPrints for research data: metadata and design, May 28, 2012 [blog post] <http://web.archive.org/web/20130331063936/http://researchdataessex.posterous.com/metadata>

