|  |  |
| --- | --- |
| **Title:** | **DataPool: Update Report for Steering Group Meeting on 31 May 2012** |
| **From:** | Mark Brown, University Librarian | **Date:** | 23 May 2012 |

# 1. Context

* The aim of the project is to build capacity within the University of Southampton to support effective data management practice. DataPool takes forward the first phase of the Roadmap identified in the previous JISC funded Institutional Data Management Blueprint IDMB)[[1]](#footnote-1)
* The project has set out five over-arching objectives:
1. Agree and implement an institutional research data policy with guidance for researchers.
2. Launch and populate an institutional data repository as part of the research data management infrastructure.
3. Develop the skills of professional support staff such as librarians and IT specialists so that they can provide a coherent institutional service to support data management.
4. Provided integrated training for PhD students through the graduate centre.
5. Embed effective data management through alignment with the work of the university’s multi-disciplinary strategic research groups.
* The work plan to take forward these objectives is divided into four work passages:
* WP1 Research Data Management System implementation
* WP2 Research Data Management policy ratification and implementation
* WP3 Integrated training and guidance and support for researchers
* WP4 Dissemination
* WP1 and WP2 were designed to run concurrently with input into WP3. In WP4 the focus currently is on the development of an institutional communication strategy.

# 2. Progress Report

* This is a brief outline of work going on under each of the work passages. At the end of each summary we have provided an indication of the key questions for the Steering Group to consider

## WP1 Research Data Management system implementation

* Work has concentrated on developing the repository model for capturing descriptive metadata and for scoping a framework for data ingest to test the viability of the high-level process architecture drawn up at the end of the IDBM project. This outlined a workflow based on a parallel development of Sharepoint and Eprints, with options for external access for the purposes of collaboration and access. This model was developed for the Sharepoint work, but informs the modelling for Eprints.



* There has been technical development work in parallel for ePrints and Sharepoint 2010 for metadata and ingest, and initial work in scoping the transfer of metadata into the repository from other University databases, for example for project profiles. Initial discussions have taken place with the Archaeology Data Centre about exporting data from the repository to the ADS; the priority is to develop workflow and guidance. This will inform the design for a SWORD2 application.
* WP1 also includes a requirements analysis to support options to make data storage and exposure more coherent and cost effective. IDMB identified a wide range of practice across the University, with the predominant model being for researchers taking responsibility themselves for storage and for the choice of media to be CD/DVD, USB/external hard-drive. Limitations to storage were met by purchasing external hard-drive or portable media.[[2]](#footnote-2) Meetings have been held with representatives from Medicine, Engineering, Life Sciences and Social Sciences, which have emphasised key requirements and vulnerabilities. Understandably this has surfaced significant differences between disciplines. Central storage was the preferred route in Maths, and was significant in Humanities and Engineering Sciences. In areas like Life Sciences there was a strong emphasis on the need for imaging data to be managed whereas in Medicine there was a very strong emphasis on control of confidential data which favoured use of non-networked, stand-alone PCs.
* The issue of storage is therefore complicated, and the team has been reviewing how best to approach the relationship between the preference for locally held storage and the need for secure storage with appropriate metadata attached. In this context cost modelling is going to be a significant factor, and the project will need to develop clarity on cost and value for money to understand the benefits of both central and distributed storage.
* Discussions have a distinction between workspace storage, with or without external access, and archival storage. There is clearly differing practice between those disciplines which organise their research workflow around the requirement to deposit ‘final’ data in a national datacentre such as those run by UKDA, NERC, ADS and ESRC. These researchers are interested in options to improve workflow, and are less likely to consider central workspace storage as crucial unless they have very large data to manage. On the other hand those researchers who are not currently under direct funder obligation to deposit in a national repository are more concerned with gaining a holistic picture of the whole data management process. Within this there is a group who are very interested in what a central storage offer could provide to systematise their workflow. One of the key factors is the new requirement by EPSRC, and the direction of policy by RCUK. Knowledge and understanding of this is as yet limited. In scoping requirements for central storage the University therefore wants us to estimate both the range and intensity of need, and the relationship between working storage and archival storage.
* Although the risks of unmanaged data are acknowledged, researchers are not easily focused on such risks. Equally there is a very mixed picture in terms of a forward commitment to open access. In some discipline areas, particularly those with national data centres, this is already accepted; the joint investigation with ADS is designed to act as a model for easing workflow. In other areas there is significant work to do to articulate pathways to access which can balance funder aspirations with significant concerns by researchers at the implications.
* IDMB set out a phased approach to data management which is now being tested in the context of a whole discipline overview, evolving external policy and the interaction with the University’s written policy.

### Issues for discussion and advice from the Steering Group (Agenda Item 3)

* Scoping storage requirements is challenging given that sector business models are as yet only evolving.
* Although researchers accept there are risks in managing data locally the costs of perceived central storage in terms of upfront investment are perceived as potentially too high. This raises the issue of the viability of a mixed mode of distributed and local storage.
* Disciplines vary in their response to the requirements for metadata; for some the creation of detailed metadata is integral to this working practice and for others this is perceived as marginal and potentially time-consuming. Many metadata schemas are comprehensive, but complex for the researcher to manage and the benefits for above minimum registry metadata are ambiguous.
* We have begun discussions with ADS on scoping the transfer of data from institutional to central repository. Technically this will be managed through SWORD, but there may be experience elsewhere which would be useful.
* Ease of interface will encourage researchers in submitting metadata. Different workflows offer different models and different levels of technical support. Is there scope for a sector approach?
* In taking forward the project to work in more detail with different disciplines, we are considering the balance between devoting effort to the setting up a complex infrastructure for specific disciplines or to develop a few indicative examples over a broad range.

## WP2 Research Data Management Policy ratification and implementation

* A consultation for the data management policy was conducted through nominated Faculty contacts and the University Research and Enterprise Advisory Group, passing to Senate, who formally adopted the policy on 29 February. The project team have been working on the guidance documents based on the requirement set out with the policy at Senate, and the policy and guidance documents are included in the agenda items. The guidance itself will be posted on the Library Data Management website pending a transfer to a more generically identified data management site. This will allow work to continue in the context of the consultation, primarily through the faculty contacts. Additional guidance already available includes information about funder requirements and links to advice from DCC.
* The process of consultation raised some interesting issues indicative of the varying views by researchers of both the significance of a policy and the nature of its requirements. Southampton has agreed a policy which is more specific and more legally expressed than the more general policies represented by other institutions such as Edinburgh. There is no doubt that the requirement of the EPSRC, and by extension RUCK policy, exercised significant influence on Senate’s decision to adopt the policy. It has been emphasised throughout the consultation that the policy is intended to be part of an iterative process, which assumes that policy and practice will evolve in response to changes in the external environment and experience internally of managing data more effectively. The major issues raised so far include the practicality of requirements for deposit and storage, the ownership of the data and licensing, ways of managing the requirement for data to be more open and monitoring of compliance.
* The project has developed a service integration model which has been influenced by presentation of a service analysis from the University of Sheffield.[[3]](#footnote-3) In developing the concept of a one stop shop to provide integrated approach to delivering institutional data management policy, the number of potential elements of service support extends across the spectrum from the institution to the local. In defining the service model it has been important to join up not only the different elements of professional support at the level of storage, data management and funder requirements, but to make a link with locally provided services at the levels of Faculties and Academic Units where there are individual staff who will have an important role in signposting individual researchers to information and support within the core services. A model of the service groupings is included in the agenda papers.

### Issues for discussion and advice from the Steering Group (Agenda Item 2)

* Our policy has deliberately been presented with clarity on responsibilities and rights for researchers, and some of what is identified may be challenging to implement.
* Most researchers are telling us that they want to keep everything forever, but the policy makes decisions on selection for post-project archiving on the basis of a ‘significance test’ lie with the PI. Is this a viable approach?
* Our policy is iterative, and the guidance is deliberately designed to be at a summary level. In preparing for launching the policy we are interested in effective ways of achieving impact.

## WP3 Integrated training guidance and support for researchers

* This work package is focused on the role of training to build expertise in data management creation and sharing within all the relevant roles in the institution, and will be developed in more detail in the second part of the project based on the initial work on infrastructure and policy. A number of pilot data management training workshops have already taken place in the areas of Archaeology and Chemistry, and a set of guidance requirements for students has been developed by Engineering, copies of which are attached to the agenda. Discussions will be taking place with the Researcher Development and Graduate Centre to include training as part of the core offering.
* A group of researchers in Engineering have produced the attached guide *Introducing Research Data* (enclosed)*.* Training based on the content has been piloted in a lecture for PGR students, and it is intended to embed this within the initial training for PGR students and to frame a course within the Graduate School as well as promoting the guide as an introduction to good practice. The DataPool Team see this as a case study which could be taken forward to other discipline areas; this would be directed at Faculties which would also have the advantage of extending understanding of the issues within academic groups.
* Initial scoping has taken place on Case Study 2 *Archiving project, tweets, blogs and websites* based on the outputs from a recent conference in Archaeology, and academic interest has indicated that the next priority should be Case Study 4 *Sharing data safely*. As a result of feedback from one of the major inter-disciplinary research areas, Life Sciences, a separate case study is being developed to look at issues relating to the data management of imaging and images which crosses between Life Sciences and Archaeology. This review is due to begin this month.

### Issues for discussion and advice from the Steering Group (Agenda Item 4)

* There is a growing corpus of material for training and development being developed at institutional level. Some generic material is not necessarily transferable into other institutional context. What priority should be given to designing generic, transferable material as against institutionally specific materials?

## WP4 Dissemination

* In terms of internal engagement the work done with Faculty representatives is seen as crucial to ensure that the process of iteration is viable and effective. The liaison model developed in the cross-service model identifies roles and remits between the services, and is designed to foster dissemination through activity and direct support rather than relying on written material alone. The aim is to raise the profile of data management in multiple fora, and to be able to blend experience across disciplines. This is particularly important in the areas such as Life Sciences where the University is promoting a strong interdisciplinary approach.
* In terms of engagement with the programme, the Team has found the event at Leeds around data management policies particularly useful and look forward to engaging with other institutions in discussions and reflections on the different ways in which policy shapes changes in understanding and culture. Regionally we would be interested in contributing to the DCC Regional Event Network as a focus for local cooperation.
* Members of the Team have been contributing to blogs and twitter, and as we have more direct work to report, this will be move from opinion and reflection to reporting developments. The Team will be presenting at RDMF9 in November.

### Issues for discussion and advice from the Steering Group (Agenda Item 5)

* The Team is interested in ways of planning and delivering internal dissemination...
* What are the key messages which we should be conveying from the Steering Group review?

Mlb.mlb.final.24.05.12

1. Institutional Data Management Blueprint, p6 <http://eprints.soton.ac.uk/196241/1/IDMB_Blueprint.pdf> [↑](#footnote-ref-1)
2. Institutional Data Management Blueprint, Initial Findings Report, p75 <http://eprints.soton.ac.uk/195155/1/idmbinitialfindingsreportv4.pdf> [↑](#footnote-ref-2)
3. See the presentation from Martin Lewis Working with other RDM actors at the recent RLUK Discussion Day on Clarifying the Roles of Libraries in Research Data Management. <http://www.rluk.ac.uk/content/clarifying-roles-libraries-research-data-management-discussion-day-find-creative-solutions> [↑](#footnote-ref-3)