|  |  |
| --- | --- |
| **Title:** | DataPool Steering Group: progress report  |
| **From:** | Wendy White | **Date:** | 7/11/12 corrected 20/11/12 |

At the last Steering Group there was a clear emphasis on the importance of supporting cultural change and identifying institutional benefits to improving research data management practice. Recent policy developments from funders have aligned parameters for the accessibility of research data to strengthening requirements for research publications. There is a focus on benefits- led activity, working with Funders and other external bodies on developing an integrated approach to improving research data management practice. The mid-phase of the project has been informed by this context as we have made progress on the key strands of project:

* Developing and rolling out service and training models to work with researchers
* Planning an evidence-based programme of support for professional services staff providing these services
* Multidisciplinary engagements
* Investigating requirements for data storage and archiving
* Testing the SharePoint and ePrints data catalogue components

### Policy engagement and integrated service support

The Research Data Management Policy was approved by Senate on 29 February 2012 and the one-stop-shop webpages with associated guidance have been launched and promoted through both formal and informal routes. There is a single email point of contact for queries, which is managed by library staff, who will then draw in other services and academic staff as needed for each enquiry. As the service develops we are working on the balance between dedicated networks of contacts for robust support and more informal knowledge of expertise within the institution. The core spine of the contact network includes the Faculty and discipline contacts in Research and Innovation Services (Research Support Officers), iSolutions (Business Relationship Managers) and Library (Academic Liaison Librarians). Enquiries have also drawn in experts with domain or discipline specific expertise e.g. data security expert, database developers with functional remit to support research, academics with technical expertise who can give advice in a multidisciplinary context.

All of the enquiries so far have been for support with developing data management plans. So far the assistance given has been for MRC plans, where the research manager had already identified a need for more support, AHRC technical appendices and first time applicants to the ESRC. We have had verbal reports via Associate Deans that the policy and guidance have been used to help complete other plans and even though they did not feel that additional support was required, the web-based information was seen as very helpful. We are working on some anonymised templates for DMPs as both academic and professional service staff would find these useful, however there is a concern that templates could lead to more of a “tick box” approach to completion and so some additional work on the “added value” of a consultation would be useful. In all cases the PI has valued help with interpreting the relevant funder requirements and we will look to further develop an approach which complements the support and guidance already provided by the funders and any relevant data archives.

Although there is no desire to “reinvent the wheel” and there is signposting to existing guidance, the process of the development of institutionally specific guidance was a helpful exercise in reflection and cross service communication. The iteration of the guidance with Faculty and service contacts was a useful learning experience and contributed to developing cross disciplinary and cross service knowledge. One of the issues for further reflection is the balance between core cross-service knowledge and the role of experts and effective referrals. <http://www.southampton.ac.uk/library/research/researchdata/>

### Training model

**PhD, Early Career Researcher and student support**

At the last Steering Group we reported on the training guide produced by Mark Scott and others which covered core elements of research data management such as file formats, version control and metadata, and then gave examples of disciplinary case studies. The feedback from the initial use of this guide for training has helped inform the development of overarching model for developing training across the institution. The core elements will form the basis of a generic first principles course which will run through our institution-wide postgraduate researcher training programme via GradBook. More tailored training will be offered on a Faculty basis and the Faculty of Humanities will work with the project on the first Faculty session. Discipline specific activity will focus on embedding content in existing programmes, lectures or events. This has been delivered in Engineering and Chemistry and also targets some masters and undergraduate students. Work with the Doctoral Training Centres will be trialled initially with the Web Science DTC. This will be seminar based and start with a discussion on issues relating to sharing data effectively. Follow up activity will be peer determined and allow for creative exploration of themes. All of these strands will run either towards the end of this year or in the first part of next semester. The feedback from these activities will iterate improvements and we should have a model that scales to support both disciplinary and multidisciplinary perspectives.

Disciplinary activity so far has involved academic staff and PhD students. For generic, Faculty and DTC training we are working with a co-delivery model involving library staff and PhD students. We will explore the balance of peer sharing, academic and professional service expertise and are open minded at this stage about involving other colleagues as the approach develops.

We also aim to contribute to the next University of Southampton Multi-Disciplinary Research Week, which this year ties in with National Science Week in March. It is hoped that this can link into a Digital Curation Centre “institutional engagement” visit where the DCC can help us work with PhD and Early Career Researchers as well as develop the skills of professional services teams. It is also an opportunity to work with data champions from the Southampton Multidisciplinary Research Forum, the network of Early Career Researchers who promote multidisciplinary excellence. The programme for last year is still available.

<http://www.southampton.ac.uk/multidisciplinary/researchweek/index.page>?

**Developing knowledge and skills in Professional Services**

A questionnaire has been developed to gather the views of staff on their confidence in a number of areas of research data management. This is in draft form and once finalised will be completed by staff across a range of professional services at Southampton and Oxford. The aim is to roll this out at Southampton before Christmas. This evidence will then inform training and development which will include the DCC institutional engagements. We also plan to select the most significant issues for two in-depth training events, with possibly one running at Southampton and one at Oxford with regional attendance. Initial briefing events and informal “research coffees” have already been run with Library staff to build engagement. Introductory briefings were run for other services before the launch of the DMP support service and we will build on this before the questionnaire roll-out so colleagues feel that this is part of a supportive process and not a criticism of weaknesses or worse an individual performance evaluation tool. We would consider running the questionnaire again in the future to benchmark improvement, but would stress that this is a whole service overview.

### Multidisciplinary engagements

The DataPool project has been committed to exploring multidisciplinary aspects of research data management and the role of early career researchers and PhD students as change agents. Funding was committed to investigate the potential contribution of staff and students with existing expertise in this category to inform and support good research data management practice. To this end Hembo Pagi (research assistant and technician) and Gareth Beale (PhD researcher) from the Faculty of Humanities have been investigating 2D and 3D imaging requirements across disciplines which will include outputs using a variety of hardware including microscopes and photographic cameras. They are now part way through their work, which is a combination of desk research and interviews with staff, which will provide:

* Comprehensive list of all facilities and software at the University of Southampton, which are used in the production or processing of raster data
* Summary document describing workflows of raster data producers/users from point of data capture/acquisition through to point of archiving
* Advisory documentation detailing best practice to raster data producers/users at the University of Southampton.

This has included consultation with the University Strategic Research Groups (USRGs) e.g. Computational Imaging; external data management and projects e.g. the Archaeology Data Service, the IT Innovation Centre and the Southampton-led EPSRC funded Research Facilities and Equipment Sharing Project.

It is possible that we will extend the case study on managing and sharing PhD data to include consideration of ways data can usefully be made available to examiners through the upgrade and examination process. There are a number of issues, so there needs to be full and sensitive discussion, but we hope the project can at least make recommendations. There is a pilot with the internal upgrade process at the moment. (See appendix for flow diagram)

We have tried to make links with PhD and ECR activity in all strands of the project, not just through the specific disciplinary engagement, and are interested in any practice elsewhere that we could draw on to add to our approach.

### Data catalogue testing

There is a forthcoming “development lunch” for the SharePoint deposit tool where a mixture of PhDs students and PIs from Music, Archaeology, Physics and Medicine will come together to test the deposit workflow. The SharePoint developers will be there and will work with on-the-fly improvements and take away issues that require further investigation. This will include testing how best to share data between nominated individuals or across multidisciplinary groups.

An EPrints data app was made available through the EPrints Bazaar which is currently being tested using the EduServe cloud facility. To work on further developments ePrints Soton needs a point upgrade so it can link to the Bazzar functionality. This upgrade is later than we would like because of the significant work undertaken earlier in the project timeline to merge the Electronics and Computer Science EPrints Repository with ePrints Soton. This upgrade is due at the end of this month which means that we can move the development to the main test server. There is interest from the community in sharing EPrints developments and Essex has kindly given us access so we can look in more detail at their work. Leeds would like to make a link as they are just getting underway with their EPrints analysis. Collaboration may help develop the EPrints app/s offered to the community through the Bazzar, ensuring that due credit goes to the developing institution/s for the work done.

Legal Services is working on a revised deposit licence which would cover anything deposited in ePrints and we will then investigate how this relates to data deposited at the workspace stage in SharePoint. We are reviewing how best to develop, select and implement end-user licences and would welcome reflection on this issue.

We are currently testing DataCite DOIs and assessing on-going cost commitments, with a view to signing up to the service. The National Crystallography Service will act as early adopters, which will help us develop criteria for DOI assignation.

### Storage and Archival Infrastructure

At the last Steering Group we discussed the need to demonstrate the value to researchers and to the institution of increased investment in the storage and archival infrastructure. We are been working on significantly increasing Faculty engagement with this process to make the case for the academic significance of investment. We have also been reviewing developments in comparator institutions like the 5TB/researcher offer already in place at Bristol. The DataPool project has also helped ensure that data requirements which are emerging through recent national and funder policies e.g. the new RCUK Policy on Open Access to Research Outputs, are foregrounded in discussions so that strategic decisions and subsequent planning take these issues into full account.

### Linked initiatives for next phase activity

The project has also been making links with a number of key initiatives to facilitate long term activity and support the mid-phase of the institutional 10 year roadmap, as produced by the Institutional Data Management Blueprint Project. It would be helpful to reflect on the balance between initiatives that reflect strategic priorities and seeking those opportunities that help shape strategic direction through proof of concept or facilitating leadership. How do we best develop a cohesive approach? These are some of the current activities:

* Securing robust support for the provision of LabTrove electronic lab notebook as a cloud-based service which provides a platform for continued innovation and significant embedding of use as an undergraduate, postgraduate and research aid - with costs of investment in provision in some way offset by the significantly reduced cost of ingest
* Promoting the work of the Digital Economy Strategic Research Group and the student Digital Champions demonstrating the methods and value of social media curation and interlinking. The project is currently coordinating a unified approach to twitter data capture and mining drawing together five formerly separate projects
* <http://digitaleconomy.soton.ac.uk/category/creative-digifest-2/>
* The contribution of Southampton to the Web Observatory to develop new dynamic approaches to accessing and visualising open data trends and themes <http://webscience.org/web-observatory/>
* Maximising the potential of Southampton’s leadership of the EPSRC funded IT as a Utility strand to inspire excellence.
* Working with the RCUK Digital Economy Sustainability Network + on shared interests in data exposure, creation and reuse and discussions with NEMODE about open research data and financial value.

### Dissemination activities

The communication strategy for the project has placed equal weight to internal communication and dissemination, as detailed in this briefing, and external opportunities to share our work and ideas with others. Here is a summary of external engagement:

**Delivered**

G.Earl (presentation), JISC DataPool, Computer Applications and Quantitative Methods in Archaeology, 26-19 March 2012, Southampton

D. Byatt, (presentation), Demystifying Research Data, RLUK/JIBS Meeting, SOAS 17 July 1012

G.Beale, (presentation), DataPool 3D: surveying institutional data practice, Creative DigiFest, Oct 11 2012, Southampton

W. White, D. Byatt, (presentation), Research Data Guidance: turning policy into practice, JISC MRD Programme Meeting, Oct 24-25 2012, Nottingham

S. Hitchcock, D. Byatt, W.White, (poster), DataPool: building capacity, developing skills, supporting researchers, JISC MRD Programme Meeting, Oct 24-25 2012, Nottingham

**Accepted**

S. Hitchcock (presentation), To architect or engineer? Lessons from DataPool on building RDM repositories, RDMF9, Nov 14-15 2012, Cambridge

W. White, D. Byatt, S. Hitchcock, M. Scott, S. Cox, G. Earl, (poster), 14-16 Jan 8th International Digital Curation Conference, Amsterdam

M. Scott, R. Boardman, P. Reed, S. Cox, Research Data Management Education for Future Curators (paper), 14-16 Jan 8th International Digital Curation Conference, Amsterdam

**Submitted**

G. Earl, G. Beale,H. Pagi, W. White, (paper) The relationship between institutional and discipline based repositories, World Archaeological Congress, Jan 13-18 2013, Jordan

**Planning**

S.Hodson etal, (UK JISCMRD/DCC Panel), Research Data Access and Preservation Summit, April 4-5 2013, Baltimore

To supplement existing internal communication channels and embedded briefings, the Digital Economy USRG is organising a spring DataPool workshop attended by representatives of all USRGs with the ECR representatives from all Faculties. Focus will be on perceived data management expertise and needs within each domain, and a cascade of DataPool information and support.

### Issues for discussion and advice from the Steering Group

What advice do you have, particularly from a disciplinary data archive perspective, on the balance between core knowledge and individual staff expertise as part of the overall service offer?

What do you consider to be the most important indicators of impact, both for the project and for improved data management practice in general?

Scoping storage and archival requirements and assessing options remain a sector-wise challenge and is a significant issue for the Russell Universities Group of IT Directors and other IT HE fora. Is there scope for shared services? At the recent programme meeting in Nottingham there seemed to be general acceptance for a mixed institutional approach with use of cloud, local data centre and other external services.

What are the critical success factors for embedding the DataPool developments into mature embedded institutional practice for the mid-phase of the roadmap?

How can we best build on PhD and Early Career Researcher engagement? Are there examples of activity elsewhere that we should be aware of?