|  |  |
| --- | --- |
| **Title:** | DataPool Steering Group: progress report  |
| **From:** | Wendy White | **Date:** | 12/03/13 |

The final project report will be available shortly, so we will take the opportunity with this final Steering Group update to highlight key areas of activity over the last period and illustrate themes for sustainability as we approach the mid-term of the 10 year roadmap. As the project draws to a close with all strands now ready for full institutional embedding, this next phase can be characterised by some principles which the project activity has helped clarify:

* Commitment to an on-going iterative approach to achieve “innovation to service” model of delivery
* Continued evidence-gathering to ensure longer-term benefits are disseminated to inform institutional decision-making and sector-wide initiatives
* Collaborative working as an underpinning requirement for success – an academic and service partnership is further strengthened by multi-disciplinary and cross-service perspectives.

Almost all aspects of the project have required collaboration. This has been true of the development of policy, joint work to review storage and investment priorities, training design and delivery, research data management planning and support services and iterating technical developments. To help support on-going collaboration in the next phase of the roadmap some of the responsibilities as PI of this project are now formally embedded in my role to continue to take initiatives forward and lead co-ordination of services. This reflects an institutional commitment to ensure that responsibility for research data management is reflected in a range of existing roles and not handled as adjunct activity.

The final suite of case studies, applications and services outlined here all embrace these three core “embedding” principles and hopefully provide a sound platform for full institutional roll-out and long term cultural change.

### Supporting data citation

The testing of institutional use of the DataCite service, through the agency of the British Library, is a good example of the “innovation to service” model. Up to now the DataCite service has mainly been used by disciplinary Data Centres or publishers, so the British Library are currently reviewing some queries about the contract which we hope they will also find helpful as the context in which DataCite is implemented expands. Once we have signed to use the live service Crystallography will in the first instance use the service and help inform the development of institutional guidance as to when we will assign Southampton DOIs to data and at what level of granularity. As part of the project the test service has been used to develop tools which will aid DOI creation. There are two tools developed and the third will be ready at project end.

1. Modification of the current code in the eCrystals repository, which mints CrossRef DOIs. This is an independent set of scripts that run on top of Eprints code, because of the nature of Eprints embargoes where these are in place. Takes appropriate metadata, registers a DOI through the DataCite API and displays the DOI on the record in the repository - included in the OAI-PMH metadata.
2. EPrints plug-in that integrates into the EPrints workflow. Takes appropriate metadata, registers a DOI, displays the DOI on the record jump-off page and includes in the OAI-PMH metadata. (Available through EPrints Bazaar)
3. Exports a selection of LabTrove Electronic Lab Notebook records as static HTML, generates appropriate metadata, assigns and registers DOI, generates "presentation quality" view with appropriate stylesheeting and title/metadata/DOI display, serves up as page associated with the originating Trove. At project end there will be a demonstrator service with the aim that additional funds will make it robust enough to go into the general release codebase , which could be part of the wider launch of LabTrove as a fully scoped cloud-based service with third party support. (DOI tools summary, Simon Coles)

### Case studies: facilitating data creation, sharing and impact

**Imaging**

The strength of the investigation of 2D and 3D imaging requirements has been the reach across disciplines and multi-disciplinary perspective. The combination of desk research and interviews with staff has provided a rich evidence base. Key findings include:

* Lack of knowledge of existing facilities
* Many examples of innovative practice, but problem solving focussed on individuals and smaller research groups
* A rapid expansion (projected to continue) in the creation and use of 3D data driven by investment in new equipment

Outputs include:

* Comprehensive list of all facilities and software at the University of Southampton, which are used in the production or processing of raster data – these will be added where relevant to the open data equipment listing from the Southampton-led EPSRC funded Research Facilities and Equipment Sharing Project. There is already evidence that improved knowledge of equipment has resulted in better use and thus value for money.
* Summary document describing workflows of 3D and raster data producers/users from point of data capture/acquisition through to point of archiving
* Advisory documentation detailing best practice to 3D and raster data producers/users at the University of Southampton
* Recommendations for improved communication across disciplines for sharing solutions and facilitating a multidisciplinary approach to problem-solving and creativity

**Impact**

This is an example of academic and service partnership with the Library, iSolutions and the Centre for Population Change collaborating to scope, build and make public migration data in a way that allows the data to be queried as an interactive tool and presented with contextual information. The use of the database will be evaluated for long term user engagement and impact. This work along with feedback from Data Management Planning support, shows that demand for web-based visualisation and manipulation tools will increase and there are lessons to be learned in terms of resource to support this and compatibility of platforms and services.

**Twitter sharing and archiving**

This has also emerged from a partnership involving University Strategic Research Groups, academic projects in digital humanities, Centre for Innovation in Technologies and Education, Student Digital Champions and an agile development approach from iSolutions. The result has been the launch of a twitter harvesting service using EPrints, for institutional research purposes in line with Twitter terms and conditions. This has already provided an archiving service for large scale conferences and for key individual and group accounts, acting as a rich resource base for analysis.

### Training

By the end of March the full training model outlined at the previous Steering Group meeting will have been delivered and tested across a range of disciplines and target groups. The feedback from these sessions will help with iterative development and has involved a partnership approach with staff from the Library, PhD researchers and academic staff all involved. The training approach and material is evolving, with the “Introducing Research Data” guide, previously circulated, as a core resource <http://eprints.soton.ac.uk/338816/>

Focus on Data Management Planning - PIs – Medicine and Health Sciences

First principles for Research Data Management - PhD – Researcher Development Graduate Centre (multidisciplinary)

Introduction to Research Data Management and case study exemplars – research staff and PhD researchers -WebScience DTC

Introduction to Research Data Management for Humanities – Early Career Researchers and PhD - Faculty of Humanities

Introduction to Research Data Management - Professional Services staff developing support services

Focus on Data Management Planning - Professional Services staff developing support services

The questionnaire on confidence with areas of research data management was completed by staff across the Library, iSolutions and Research and Innovation Services. The analysis is just being completed and will provide evidence for targeted development of future training. The questionnaire is also being deployed at Oxford which will give us comparative information.

### Data catalogue and archiving

The EPrints data app developed by Essex has been slightly modified to fit with the Southampton workflow and is now being tested. A useful conference call between interested EPrints data users (Essex, Glasgow, Leeds, Southampton, EPrints Services) arranged by Louise Corti should help keep core principles across all sites, thus supporting inter-operability and adherence to relevant standards. The plan is for a soft launch and to take an iterative approach to minor development as take up of the service expands. There will be an additional 20TB of storage available to support the enhanced service.

The project is also working with Arkivum on a proof of concept link that integrates Arkivum’s A-Stor archive service with the ePrints repository at Southampton. A number of use cases will be developed for the storage and access of research data using the Arkivum service. One or more of the use cases will be implemented in a demonstrator to show how they would work in practice. Likely candidates are:

* Movement of research data into archive storage (Arkivum service) for data that hasn’t been accessed for a given period, or data that is not expected to be accessed frequently
* Retention scheduling and review of research data according to University policy, for example being able to set a retention period that subsequently triggers a review of whether the data should be
* destroyed, retained or made public. (from Statement of Work, Matthew Addis)

### Next phase governance and sustainability

There are some points that the project team would like to emphasise for discussion and steer.

1. There will be a paper tabled for the next Research and Enterprise Advisory Group meeting on options for phased investment based on consultation with Associate Deans Research. This will also propose continuation of a reconstituted Steering Group reporting to REAG to provide continued steer for the embedding of research data management services and related policy development.
2. The SharePoint data development work is available but dependent on longer term institutional plan for SharePoint. We suggest revisiting the strategic view of VRE options, given known demand for managing multidisciplinary and multi-stakeholder access to versions of working data.
3. Commitment to assessing uptake of data catalogue once launched and exploring link between service development and policy implementation.
4. Important to work closely with Researcher Development Graduate Centre, Faculties and DTCs on full embedding of training to support long term development of next generation researchers.
5. Use launch of LabTrove as open source and third-party supported cloud-based service as exemplar for on-going “innovation to service” model.
6. Continue to seek sector-wide links to data initiatives e.g. through EPSRC IT as a Utility strand, Digital Humanities, open data and big data developments.

 **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 since the last Steering Group:

**Delivered**

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

D. Byatt, G. Beale, H. Pagi, S. Hitchcock, M. Scott, S. Cox, G. Earl, W. White (poster), Working Collaboratively with PhD and Early Career Researchers: agents for change, 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

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

**Accepted**

Gareth Beale, Richard Boardman, Graeme Earl, Steve Hitchcock, and Hembo Pagi (interactive paper),Supporting Data Management for 3D andRaster Data: A Case Study, Society for Imaging Science and Technology Archiving, April 2-5 2013, Washington D.C

Gareth Beale, Imaging work looking at multi-disciplinary requirements and use of equipment**,** Jisc Managing Research Data Programme Workshop: Achievements, Challenges and Recommendations, Aston Business School, 25-26 March 2013

Wendy White, Implementing DOIs for Data, Jisc Managing Research Data Programme Workshop: Achievements, Challenges and Recommendations, Aston Business School, 25-26 March 2013