University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

JISC

JISC Final Report

Project Information			
Project Identifier	*To be completed by JISC*		
Project Title	Research360		
Project Hashtag	#research360		
Start Date	1st October 2011	**End Date**	31st July 2013
Lead Institution	University of Bath		
Project Director	Professor Matthew Davidson Dr Liz Lyon		
Project Manager	Dr Catherine Pink		
Contact email	research-data@bath.ac.uk		
Partner Institutions	Digital Curation Centre, UKOLN		
Project Web URL	http://blogs.bath.ac.uk/research360 http://www.ukoln.ac.uk/projects/research360/		
Programme Name	Managing Research Data Programme 2011-13		
Programme Manager	Dr Simon Hodson		

Document Information			
Author(s)	Dr Catherine Pink		
Project Role(s)	Data Scientist and Project Manager		
Date	June 2013	**Filename**	R360_JiscFinalReport.doc
URL	NA		
Access	This report is for general dissemination		

Document History		
Version	**Date**	**Comments**
v0.1	15/05/2013	First draft by C.J.Pink.
v0.2	28/05/2013	Updated by C.J.Pink to incorporate comments from project directors.
v0.3	09/07/2013	Updated by C.J.Pink to incorporate comments from Research360 Steering Group and Research360 project team.
v1.0	22/07/2013	Final draft incorporating Steering Group comments on v0.3.

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

Acknowledgements

1   Project Summary

The Research360 project aimed to develop human and technical infrastructure to facilitate management of the University of Bath's research data, in response to increasing external drivers for access to publicly funded research outputs.

The project successfully laid the foundations for an embedded data management service. This included a roadmap setting out how compliance with the EPSRC's expectations for research data will be achieved. To support the implementation of this Roadmap for EPSRC, a new, high-level Research Data Sustainability Group was established.

The project created two new data management roles within the University and, by delivery of a successful draft Business Case, succeeded in permanently retaining these essential, central roles. Supporting this draft Business Case was a range of guidance, aimed at enabling other institutions to develop similar business cases. This guidance demonstrated how best practice in research data management would benefit the institution, external collaborators and society and also how it would contribute to the Research Excellence Framework (REF) 2014. The project found that responding to researcher demand for training, storage and data publication was a strong motivator of continuing investment in data management activities.

Through an institution-wide survey, the project identified a lack of central policy and co-ordination, and difficulties storing and sharing research active data, as main priorities for the University to address. Research360 therefore developed a draft data management policy for the University and documented the process in a set of guidelines aimed at the wider community. The project redesigned the University's data management website, as a focal point for researchers searching for help. It also developed specific guidance on the storage of research active data. This guidance was supported by the development of a new training workshop and an online training module, designed to give researchers an introduction to the issues and responsibilities of research data management.

Research360 explored integrating the Sakai virtual research environment with the SWORD2 deposit protocol and reported the experience and recommendations to the Sakai community. The project also

---

[1] Centre for Sustainable Chemical Technologies: http://www.bath.ac.uk/csct/

worked with EPrints Services to develop a pilot institutional data repository, which will facilitate cataloguing and publishing research data.

Through a broad dissemination programme and close coordination with a range of stakeholders across the University, the Research360 project team established a network of local, national and international advisors, who will prove essential partners as the ongoing data management service continues to develop.

The project concluded that establishment of a pilot data management service requires considerable resource and that many aspects of data management require further investigation. As such, the project strongly recommends that Jisc continue to support programmes and activities in research data management.

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)
Requirements Report (Deliverable 2.1)	Report detailing the results of the University of Bath's 2012 Research Data Survey, a series of case studies and a management view of current data management practice across the institution. The report has a particular focus on collaborative research between the University of Bath and commercial partners. Redacted version will be available from: http://opus.bath.ac.uk/36361/.
CARDIO feedback report (Work Package 2)	Report capturing the project's experience using the DCC's CARDIO for surveying the institution's current research data status. This included feedback from a workshop where CARDIO was tested by librarians. Report submitted to the DCC but will not be published externally.
Bristol Online Survey template (Work Package 2)	Questionnaire in Bristol Online Surveys, used to capture the results reported on for the Requirements Report. Full template can be viewed at https://www.survey.bath.ac.uk/research_data_template [accessed 24/07/2013]. Institutions wishing to use the template can contact the Research360 project manager, who will then share the survey with the institutions Bristol Online Survey account. The full questionnaire is also included as an appendix to the Requirement Report.
University of Bath Roadmap for EPSRC (Deliverable 3.1)	Report setting out the current institutional status relative to the EPSRC's 9 expectations for research data and sets out 21 objectives and activities for how the University will achieve

	compliance with the expectations by the EPSRC's May 2015 deadline. Redacted version available from: http://opus.bath.ac.uk/31279/ [accessed 09/07/2013].
Research360 draft Business Case to Support Future Research Data Management in the University (Deliverable 3.2)	Confidential, internal document setting out the case for ongoing investment in research data management, listing the benefits of good data management to the institution and the risks of not doing so. The draft business case identified potential sources of funding and presented the anticipated service that would be delivered from four different investment levels: PreR360 status, minimal, optimal and exemplary. It also made 8 recommendations to the University that, if implemented, would minimise the current shortfall in current data management support. Internal, confidential report not due for publication. No redacted version is planned for publication.
Benefits from Research Data Management in Universities for Industry and Not-for-Profit Research Partners (Task 3.2)	Document produced in collaboration with Charles Beagrie Ltd and identifing how different internal and external stakeholder groups, including commercial and not-for-profit research collaborators, would benefit from good research data management. Full version available from: http://opus.bath.ac.uk/32509/ [accessed 09/07/2013].
Research Data Management and REF 2014 (Task 3.2)	Guidelines produced in collaboration with Charles Beagrie Ltd, providing a brief overview of how research data and research data management are referenced in the REF 2014 guidance documents and can contribute to the three REF 2014 elements. Full version retained as confidential to the University. Redacted version available from: http://opus.bath.ac.uk/35518/ [accessed 09/07/2013].
DMPonline evaluation report (Deliverable 3.3)	Report on the suitability of the DMPonline tool for postgraduate research students based in Doctoral Training Centres. Full version available from: http://opus.bath.ac.uk/36206/ [accessed 30/07/2013].
Institutional DMP template (Deliverable 3.3)	DMP template for the University of Bath. A version of the template with suggested answers is also available. Full version of both templates available from http://opus.bath.ac.uk/36360/ [accessed 30/07/2013].
Institutional DMP guidance (Deliverable 3.3)	Guidance for the completion of the University of Bath DMP templates. Full version available from: http://opus.bath.ac.uk/36359/ [accessed 30/07/2013].

Data Management Plan for Postgraduate Researchers (Deliverable 3.4)	Data Management Planning template designed for use by postgraduate research students, once a project has been started. Full version available from: http://opus.bath.ac.uk/30772/ [accessed 09/05/2013].
DMP DTC Guidelines (Deliverable 3.4)	Guidelines for completion of the Data Management Plan for Postgraduate Researchers, aimed particularly at postgraduate students based in Doctoral Training Centres. Full version available from: http://opus.bath.ac.uk/36009/ [accessed 30/07/2013].
Policy Development Guidelines (Deliverable 4.1)	In development but not due for completion until after the project has finished. Full version will be published via Opus.
Faculty-Industry RDM Case Studies for Policy Development (Deliverable 4.2)	In development but not due for completion until after the project has finished. Full version will be published via Opus.
University of Bath Data Management Policy (Deliverable 4.3)	In development, due for consideration in the next stage of the approval process during October 2013. Full version of the policy will be published via the University of Bath website.
Sakai-SWORD2 Integration Development Report (Deliverable 5.2)	Report documenting relevant experiences encountered when developing an integration between the CLE version of the Sakai virtual research environment (VRE) and the SWORD2 deposit specification, with the intention of integrating data deposit seamlessly into the research workflow. The report also makes recommendations for future developers seeking to build on the initial work achieved. Full report available from: http://opus.bath.ac.uk/35540/ [accessed 06/06/2013].
Recommendations for Improvements to Sakai Documentation (Deliverable 5.2)	Report containing a list of recommendations for the Sakai development community that, if implemented, are likely to improve the experience of future developers who are new to Sakai. The recommendations are listed in order of how easy it is anticipated they would be to implement. Full version available from: http://opus.bath.ac.uk/35541/ [accessed 06/06/2013].
Patch file for Sakai (Deliverable 5.2)	Software patch file for Sakai, affecting the content tool (also known as the resources tool) module and the internationalisation information. Unfinished code that does not compile, resulting from the terminated attempt to develop and integration between Sakai and SWORD2.

	Available from https://bitbucket.org/research360/sakai-sword-patch/src [accessed 06/06/2013].
EPrints Integration with the Hitachi Content Platform - Report (Deliverable 5.2)	Report documenting development of a proof-of-concept integration between EPrints and the University of Bath's Hitachi Content Platform (HCP) next-generation storage solution. The aim of the work was to determine whether the HCP could be used to manage archived research data. Full report available from http://opus.bath.ac.uk/35532/ [accessed 06/06/2013].
EPrints-Hitachi Content Platform plugin and associated 'readme' documentation (Deliverable 5.2)	Perl file containing the code required for the proof-of-concept integration between EPrints and the Hitachi Content Platform. The software is accompanied by a 'readme' text file containing installation instructions. Available from: https://bitbucket.org/research360/eprints-hcp [accessed 06/06/2013].
Institutional Data Repository User Stories (Deliverable 5.2)	Document setting out the requirements as a University for how a new institutional data repository should work. User stories were captured from interviews with key stakeholders and from the results of the 2013 Research Data survey. Full version available from http://opus.bath.ac.uk/34082/ [accessed 09/05/2013].
Pilot Institutional Data Repository (Deliverable 5.3)	Pilot research data repository for the University of Bath, developed by EPrints Services, to facilitate sharing and publication of research data. In development, initial pilot due for completion during August 2013. The University of Bath Research Data Archive will be available from: researchdata.bath.ac.uk
Data Storage Guidelines (Deliverable 5.3)	Guidance information on current University services to support the storage and sharing of research active data. Full 'Storage Guidelines' available from: http://www.bath.ac.uk/bucs/aboutbucs/policies-guidelines/guidelines_storage.html [accessed 09/05/2013]. This guidance is supported by additional guidance developed by Bath University Computing Services (BUCS) outside Research360, on the use of cloud storage for University information: Full 'Advice on storing University of Bath information in the cloud' available from: http://www.bath.ac.uk/bucs/aboutbucs/policies-guidelines/cloud_storing_data_guidelines.html [accessed 09/05/2013].

RIM Guidelines for Data Driven Reporting and Institutional Benchmarking (Deliverable 5.4)	Report summarising 4 key areas in which research data is or could be included in institutional reporting, including the Research Excellence Framework, Research Outputs System, Pure Research Information Management System and institutional benchmarking. The report also makes recommendations for how these reporting processes could be improved to facilitate data driven reporting. In development but not due for completion until after the project has finished. Full version will be published via Opus.
VRE RDM Training Module (Deliverable 6.1)	Online data management training module, developed in Xerte and to be embedded in Moodle, the University's Virtual Learning Environment. The training module utilises material from a range of existing training resources and presents a short, high level overview of the key elements of research data management. In development but not due for completion until after the project has finished. Full version will be published as an Open Educational Resource via Jorum.
Managing Your Research Data 'PG skills' training workshop (Deliverable 6.2)	Face-to-face training workshop for postgraduate research students, aimed to give a 2 hour introduction to research data management. Slides available from: http://opus.bath.ac.uk/32296/ [accessed 09/05/2013].
University of Bath Research Data Management Website (Deliverable 6.2)	Re-designed University of Bath Data Management website to act as a focal point for information and guidance on data management, for both internal and external communities. Available at: http://www.bath.ac.uk/research/data [accessed 09/05/2013].
Dissemination Programme (Deliverable 7.1)	Programme of attendance at relevant events, conferences, workshops and meetings where experiences from the project were shared with the wider research data management community. These events included: - All DataCite workshops run at the British Library, including presenting an invited talk on working with commercial partners; - All Jisc MRD Programme events; - Establishing and building close relationships with regional universities on the Jisc MRD programme: Open Exeter, data.bris and 'A Pilot Study in the Health and Life Sciences'. Research360 set up the first of a number of regional meetings for this group;

	• A workshop, presentation and poster at the IDCC13 conference;
	• Presentation of a poster at the ARMA 2012 conference;
	• Running a workshop on using research data in teaching skills at the ALT-C 2012 conference;
	• Contributing to the DCC Institutional Engagement programme by advising the University of Warwick on developing the *Roadmap for EPSRC*;
	• Advising Cardiff University on the project's experiences of policy development;
	• An international programme of keynote presentations.
	Selected presentations and posters from these dissemination activities are available from these sources:
	Catherine Pink's publications: http://opus.bath.ac.uk/view/person_id/3237.html [accessed 10/05/2013].
	Jez Cope's publications: http://opus.bath.ac.uk/view/person_id/5476.html [accessed 10/05/2013].
	Liz Lyon's presentations: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/presentations.html [accessed 28/05/2013].
	UKOLN Research360 project page: http://www.ukoln.ac.uk/projects/research360/ [accessed 04/07/2013].
	Neil Beagrie's presentations relating to Work Package 3.2: http://www.beagrie.com/publications.php?t=talks [accessed 04/07/2013].

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

Project Aims

The Research360 project aimed to develop human and technical infrastructure for research data management at the University of Bath, focussing on all aspects of the research lifecycle, from project planning to data capture, deposit and reuse. The project originally intended to use a transferrable Faculty cascade approach, focussing on the Faculties of Science and Engineering, but with the Research Data Management (RDM) implementation designed to scale across the whole institution. Two key aims for the project were to meet the EPSRC Policy Framework on Research Data, an important research funder for the University, and to explore particular data management issues arising from research collaborations with industry. The close association with industry and commerce is written into the University of Bath's Charter of Incorporation (University of Bath, 1966) and this aspect of research was therefore intended to be a unique aspect of the Research360 project.

Original Project Team

In order to achieve these original objectives, the project established two new posts. A new institutional Data Scientist was recruited to interface with existing support services, including the Library, BUCS, the Research Development and Support Office (RDSO) and Faculties. Unlike existing staff members, the sole focus of the Data Scientist's role within the institution was RDM activities. A Technical Data

Co-ordinator was seconded to the project on a full time basis, from their existing role within a cross-Faculty Doctoral Training Centre (DTC). The Technical Data Co-ordinator's role was to provide general research technology expertise on data repository development, virtual research environments and electronic lab notebooks, and to provide specific co-ordination and communication with technical services including BUCS. The DTC in which the Technical Data Co-ordinator was based provided a central hub for data management planning and training activities.

These two posts were supported by a project team distributed across key professional service departments, who contributed to individual tasks or work packages. Within this team would be a designated project manager, seconded on a part time basis from RDSO. The team were assisted by a number of external consultants, brought in to apply their expertise to the development of key work packages.

Expanded Project Focus

Although the project originally intended to focus on the Faculties of Science and Engineering, it rapidly became apparent that there was interest in RDM from across the University. More requests for data management advice, via the newly established research-data@bath.ac.uk contact address, were received from outside the Faculties of Science and Engineering than within them (Figure 1). Similarly, when registration for the first few RDM training workshops was opened up to all research postgraduates, a substantial proportion of attendees came from outside the focal Faculties (Figure 2).
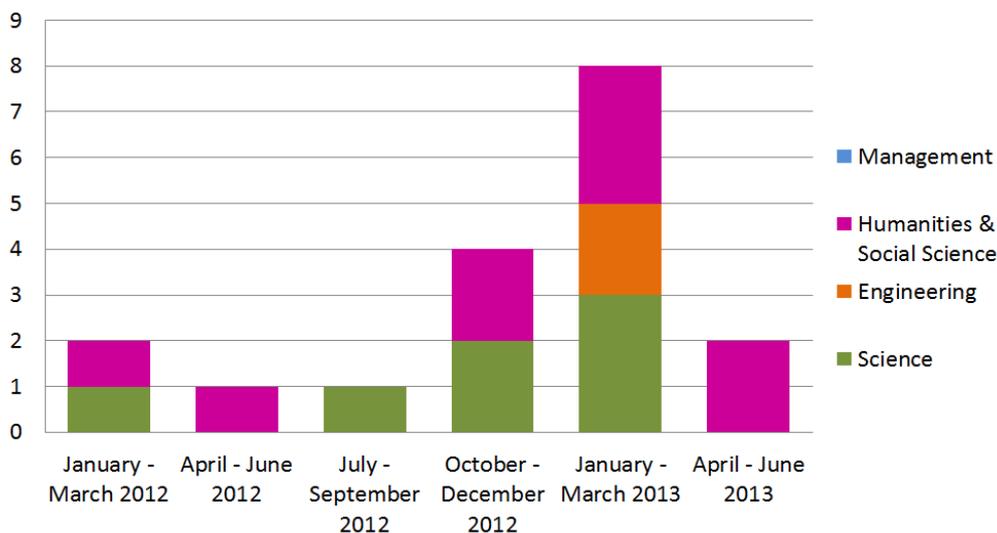

Figure 1: More individual requests for data management support were received from outside the focal Faculties of Science and Engineering than within.

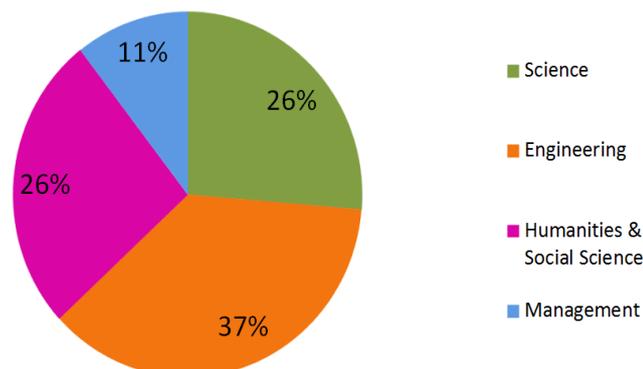**Proportion of RDM training workshop attendees
from each Faculty**

Figure 2: Voluntary attendees for research data management training workshops came from all University Faculties.

Both the 2012 Data Asset Framework (Jones, 2012) and the Research Data Survey (Pink et al. 2013, unpublished) had highlighted a considerable amount of uncertainty about data management and a lack of awareness of, and trust in, the existing University infrastructure to support the data management process. Declining support to researchers who had requested help would have reinforced these misconceptions. In order to build the necessary cultural change across the University to ensure compliance with EPSRC expectations, it was considered vital to engage and successfully help as many researchers as possible, so that the reputation and status of research data management would be enhanced. It was therefore decided to expand the focus of the project to the entire University. This was ultimately found to be beneficial as it uncovered a broader diversity of data management requirements that the developing infrastructure would have to cater for. Such requirements included interview based data and data protection issues as well as ESRC and AHRC data management planning requirements.

The project had originally intended to explore data management issues arising from collaborations with industry. Although two key aspects of data management were identified – policy and sharing of research active data – more detailed investigation was sacrificed in order to meet the more generic needs of the University. Further, difficulty in data sharing during active projects was found to be an issue for all researchers, whether sharing data within their project teams or with external partners, both commercial and academic. It was felt that providing a complete solution would be a distinct project that fell outside the scope and resource of the Research360 project.

<u>Evolving Project Responsibilities</u>

Another aspect of the Research360's approach that was revised during the project related to the team assigned to the project. It became rapidly apparent that the size, complexity and scope of the Research360 project required considerably more than the anticipated 0.2FTE to effectively manage. Day-to-day management and co-ordination of project activities required close involvement and specialist data management knowledge. As such, project management responsibilities were taken over by the Data Scientist. This allowed for stronger inter-relationships to be created between different work packages and for project outputs to respond to and benefit from the rapidly developing external data management environment. However, whilst overall this additional responsibility was advantageous, it also came at a cost to the project: less time was available for the Data Scientist to work on project outputs, so it became necessary to prioritise delivery of those aspects of the project that would both provide the basis of the infrastructure necessary for EPSRC compliance, and

establish the support and resource necessary to continue providing the new RDM service once the project had finished.

Project Methodologies

An important aspect of the approach taken to delivering project outputs was to build on existing resources wherever possible. For example, the *University of Bath Roadmap for EPSRC* was initially based on Monash University's "Research Data Management Strategy and Strategic Plan 2012-2015" (Beitz et al, 2012), which was re-written to align with the University of Bath's strategic aims and then mapped to the EPSRC's 9 policy expectations. Similarly, training resources for face-to-face workshops were based on material originally generated by the Jisc-funded MANTRA[2] and DataTrain[3] projects. The project also utilised the work and experiences of other Jisc-funded projects on the same MRD 2011-2013 programme. For example, the design of the research data survey was initially based on the themes from the CARDIO survey, but informed by the questionnaire published by the Open Exeter project (Open Exeter Project Team, 2102).

A number of approaches were used to gather the information required for the Business Case. A consultant was employed to develop the business case and he conducted a series of semi-structured interviews with key stakeholders across the University, including a number of heads of professional service departments. These interviews provided insight into the current data management position of the institution, including current and future funding streams. The Keeping Research Data Safe[4] (KRDS) framework was applied to both internal and external stakeholders, including industrial and not-for-profit research partners, in order to identify benefits from improved research data management. To provide an insight into the external data management environment, with particular reference to how other peer-group institutions were tackling similar challenges, members of the project team attended RDM community events such as workshops and conferences, where relevant topics were discussed. Finally, the consultant, project manager and project directors worked closely to develop a series of RDM service levels that could be achieved with increasing levels of investment.

A similar approach was applied to the development of technical infrastructure. The specification for the development of a new pilot institutional data repository consisted of a series of features that a University of Bath data repository 'must', 'should', 'could' and 'won't (but might)' have. These features were developed from a series of data repository user stories[5], which were captured from interviews with key stakeholders and from the results of the 2013 research data survey. Delivery of this infrastructure required support from all institutional stakeholders, which was achieved through regular communication and discussion, facilitated by the project manager.

Another approach taken to developing project outputs was collaborative work by sub-groups within the project team. For example, both drafting the new research data policy and content development for the new online training module were undertaken over a series meetings by groups of 'experts' working around a single computer on one document. The advantage of this approach meant that key issues could be discussed and resolved without protracted email discussions or versioning issues. However, for policy development it resulted in excessive focus on and re-writing of some aspects of the policy, whilst the overall aim and tone of the policy was lost. As such, a second major drafting of the policy was taken over by the Data Scientist, who had oversight of all relevant data management issues relating to the policy.

The final approach taken to development of Research360 project outputs was regular consultation and feedback, both within the project team and with key stakeholders. For example, the training workshop was regularly updated in light of feedback from participants and both the training workshop

---

[2] MANTRA training course: http://datalib.edina.ac.uk/mantra/index.html [accessed 13/05/2013]

[3] DataTrain resources: http://archaeologydataservice.ac.uk/learning/DataTrain [accessed 13/05/2013].

[4] Keeping Research Data Safe: http://www.beagrie.com/krds.php [accessed 04/07/2013].

[5] Institutional Data Repository User Stories: http://opus.bath.ac.uk/34082/ [accessed 09/05/2013].

and data management planning templates were initially tested in the focal DTC. Industry stakeholders provided feedback on the 'Benefits from Research Data Management in Universities for Industry and Not-for-Profit Research Partners' and other Jisc-funded projects provided feedback on the 'Sakai-SWORD2 Integration Development Report'.

2.3 What did you learn?

The project originally anticipated that external policies would be the strongest driver of investment in data management. The project's focus was initially on compliance with the EPSRC Policy Framework on Research Data, with both the *University of Bath Roadmap for EPSRC* and a substantial proportion of the draft *RDM Business Case* centred on the current and future status of the University with respect to the EPSRC's expectations. Both documents drew attention to the risk of losing a large component of the University's research income if compliance was not achieved by May 2015.

However, the project discovered that the ability to demonstrate demand from researchers was also an effective driver of institutional data management activities. For example, a key aspect of both the *University of Bath Roadmap for EPSRC* and the draft *RDM Business Case* was the need for an institutional data repository to achieve full compliance with the EPSRC expectations. However, despite the high-level endorsement of these documents, progress on development of a pilot data repository was stalled due to a lack of agreement within the project team on how best to approach the development work. The impasse was eventually broken by requests from several prominent researchers, who independently demanded an institutional data repository through which they could publish their research data. One researcher wanted to make the data supporting an article permanently accessible and, more importantly, wanted to use persistent identifiers to link to different files within the dataset so that they could be directly referenced within the text of the article. Another researcher wanted to ensure that all members of their research group could make their data outputs publicly accessible. Although they were currently using a project website to make selected data available, management of and access to this website was restricted to the PI, who also recognised that it did not fulfil funder requirements for long term data access. That demand for a data repository was primarily focused on the researchers' desire to share their data is an encouraging indicator of cultural change.

User testing of the newly developed data management training workshops helped to refine their format and content. Based on confidential feedback provided by participants at the end of workshops, the project team learnt that a hands-on workshop approach required structured exercises to develop skills as well as discussion elements. Participants tended to favour lecture-style workshops, particularly as early career researchers tend to be used to lectures and favour this format for gathering new information - although they are happy to take part in general discussions about data management, they generally did not consider this to be a worthwhile use of their time. In addition, researchers preferred content that demonstrated 'how', 'what' and 'where' rather than 'soft' skills learning. For example, positive feedback was received when workshops included: practical advice on how to scan notebooks; examples of good file names or file formats; software recommendations; and data management planning templates.

From the experience of developing the Sakai Virtual Research Environment (VRE), the Research360 project found that the substantial technical difficulties encountered were magnified by the lack of suitable documentation, which was frequently out-dated, inaccurate, missing or difficult to find. This was found to be particularly problematic for developers new to Sakai, as the current culture within the Sakai community was, to quote reviewers of the Sakai-SWORD2 Integration Development Report, *"if you think it's broke then you fix it"* or *"then write it if you need it."* However, as the report's reviewers also noted *"[t]his is fine in places where you know what you're doing, but when you don't, you don't have the knowledge to create the required resources in order to help yourself…"* and *"[t]here is a certain amount of missing 'middle rung' documentation for those that may be talented Java coders but are not au fait with the Sakai environment….this sort of documentation can ONLY be written by the*

knowledgeable, pre-existing core of the sakai community, NOT, as the culture would have it, the complainants." The Research360 project also found that Sakai installation from source can be hugely complex and time consuming. The project developer initially allocated 10 days to complete all of the development work, which was extended to over 40 days before this aspect of the development work was terminated. As a reviewer commented: *"Setting 10 days to discover how to 'install' Sakai is very brave, setting 40 days to install and understand how to develop for Sakai is still quite brave!"*

For the online training module, the collaborative group approach described in Section 3.2 did work efficiently, with development progressing according to the original schedule. However, the project team felt that another aspect of the approach that did not work so well was that the process started with subject matter experts collectively identifying required content for the module. The actual content was then created by a non-expert in data management using pointers to existing resources provided by the expert group. Although edits and revisions were made using the collaborative method previously described, this approach resulted in many high quality resources not being fully utilised. In hindsight, a more effective approach would have been for the group to systematically work through an existing resource, such as MANTRA, and extracting or modifying material as required. It is not known how long the latter approach would have taken. However, assuming that directly modifying an existing resource could have been completed by a single person over a few days, then it is estimated that this could have saved over 70hrs of time: a conservative estimate of the time spent on this work package by additional members of the project team. It would also have saved the considerable additional cost of hiring a module design expert to create a new resource almost from new.

2.4 Immediate Impact

Institutional Support for RDM

The most significant impact of the Research360 project was its success in permanently establishing a data management service for the University. The key difference made by Research360 was that, by establishing two dedicated roles, communication between key stakeholder staff and departments was considerably improved. This was critical to developing the interest in and support for data management. In addition, for the first time the University had a focal contact for data management responsibilities, activities and queries, both within the professional support services and the research community. The project also enabled the development of two key reports - the *Roadmap for EPSRC* and draft *RDM Business Case -* which were essential both to demonstrating the need for and to securing the funding for a permanent, embedded data management service. Professor Jane Millar, Pro-Vice Chancellor for Research and chair of the project's steering group commented:

> *"The business case presented a clear case for investment in research data management. The University is continuing to invest in this area. The approval of investment in new posts is a specific outcome from the project for the University. This directly contributes to the implementation of our Research Data Management Roadmap for EPSRC, in order to ensure that the University is compliant with EPSRC and other funder expectations by 2015 and beyond".*

Feedback from both academic and support staff has been positive. A senior academic, who had requested help with preparation of a data management plan and the future publication of their research data, initially commented "it's good to know that someone in the university is thinking about these issues". They also informed the project team that "the EU consultant advising me on the proposal was very impressed and it has definitely helped our case." A member of the professional support staff, outside of the project team, commented on the project, saying "this is something that is really important to the university". The project has successfully raised the profile of data management at all levels in the University, culminating in the approval of a new Research Data Sustainability Group. That this group is composed of the heads of all professional service departments,

representatives from the senior management team and from the academic community, demonstrates the University's recognition of the importance of data management.

Increased Use of Dedicated Data Storage

Like most of the projects on the Jisc MRD programme[6] and universities across the UK (Beagrie, Beagrie & Rowlands, 2009), Research360's research data survey revealed that the University's dedicated research storage facilities were not used as much as alternative, less secure or resilient types of storage such as local hard drives or USB sticks (Figure 3). When survey participants were asked why they did not use the University's research storage, the most commonly cited reason was the need for a reliable internet and VPN connection to access the storage, a perceived barrier to off campus or mobile working. However, the survey also revealed frequent misconceptions, including a perceived high cost for storage (1TB is in fact provided free to funded projects) and a perceived lack of access controls for 'shared' storage (access permissions may be restricted to named individuals within a project or research group).
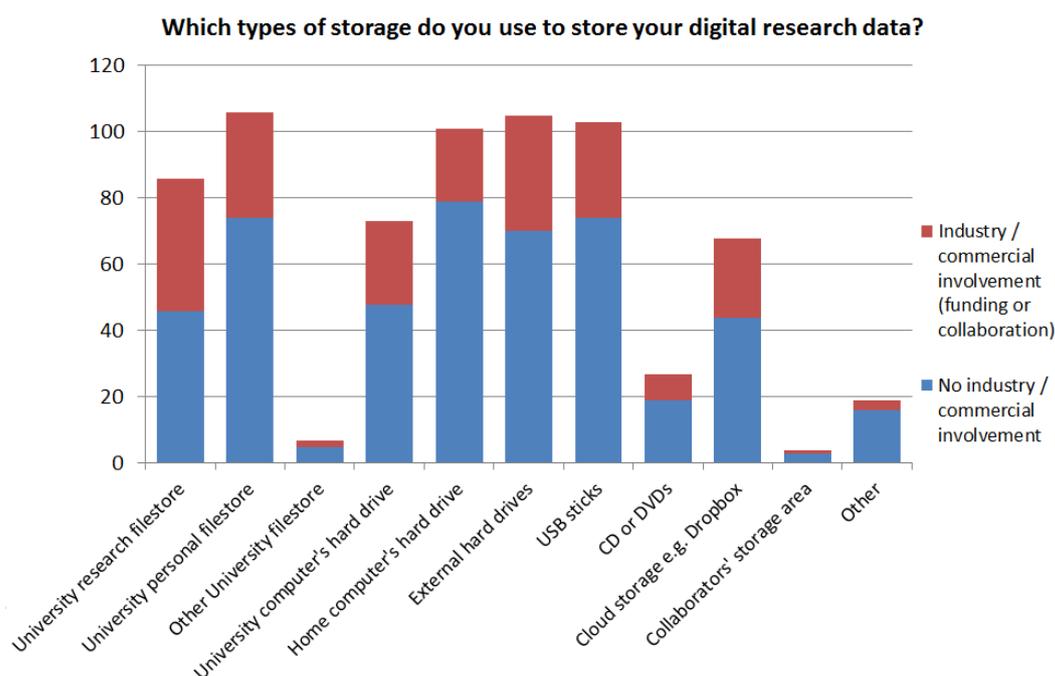

Figure 3: Survey respondents were asked to select all locations where they store research active digital data. The survey revealed that in addition to University provided research storage space, personal or local file facilities are also commonly used to store research data.

The Research360 project team therefore started promoting the provision of University research storage as the optimal storage medium for research active data. This was achieved through new RDM training workshops, the upgraded RDM website, individual requests for help and the provision of new data storage guidelines. As a result, there was an increase demand for and use of centrally managed research storage space, often coinciding with key advocacy events (Figure 4), thus improving the security and resilience of research active data and reducing the risk of loss or inappropriate release of data.

---

[6] Jisc MRD Programme 2011-2013:
http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx
[Accessed 04/07/2013]

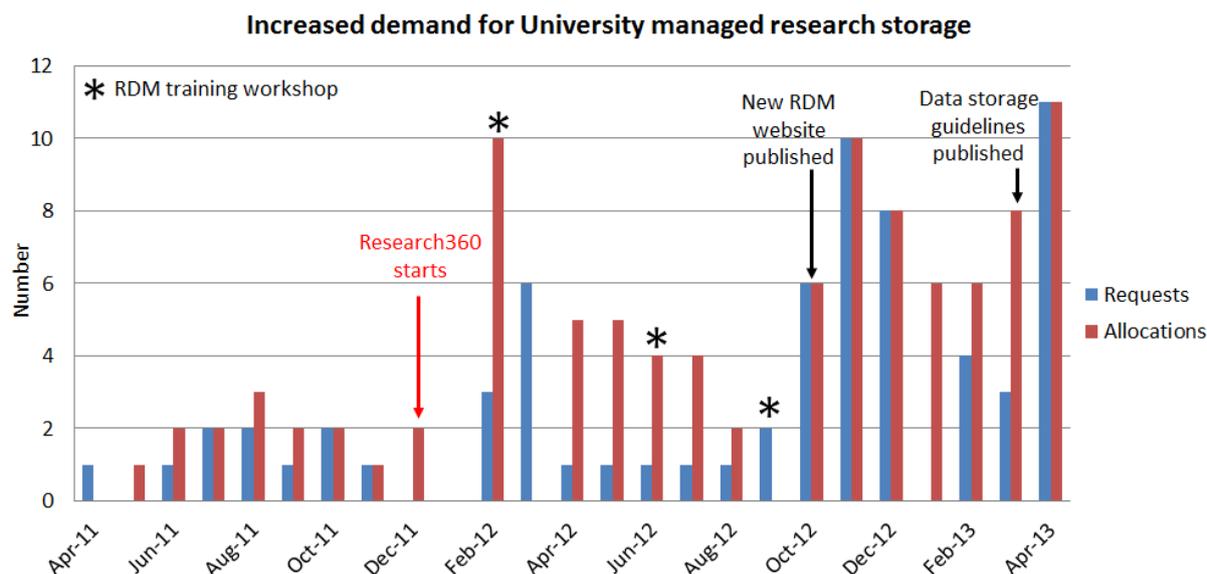**Increased demand for University managed research storage**

Figure 4: Since the start of Research, there has been an increase both in requests for storage and in the number of research projects allocated at least 1TB of University managed storage. Requests for data storage have coincided with project advocacy activities. Note that these data are likely to underestimate demand: requests are based on tickets remaining in the data storage queue and do not include tickets moved to other queues; allocations are based on storage partitions created on the newest data storage hardware and do not include use of existing data storage infrastructure.

Community Benefits

The wider data management and higher education communities have already benefited from the Research360 project. Many project outputs have already been reused: three Universities requested permission to re-use the design of the new RDM website; two Universities expressed an interest in adapting the *University of Bath Roadmap for EPSRC* for their institutions; several universities adopted the *Data Management Plan for Postgraduate Researchers*; and other Universities re-used slides from the Managing Your Research Data 'PG skills' training workshop. In addition, the Research360 project has been the subject of a Digital Curation Centre (DCC) Case Study on 'Increasing Participation in Internal RDM Training Sessions'. Members of the project team have been asked to provide advice to UK universities outside of the Jisc MRD 2011-2013 programme on aspects of both roadmap and policy development, and to international universities and a commercial company on issues related to the management of commercial research data. In addition, feedback from a student focus group on data management planning has contributed to the DCC's future plans for DMPonline.

Towards the end of the project, a number of transferable guidelines will be published. The *Benefits from Research Data Management in Universities for Industry and Not-for-Profit Research Partners* guidance, designed to assist in the development of RDM business cases has already been downloaded over 400 times and used by at least one project on the Jisc MRD 2011-2013 programme. Guidance on how research data and research data management can contribute to the three REF 2014 elements is already being used by the institution and a version was published externally in June 2013. Guidelines on data driven reporting and institutional benchmarking are currently in development, due for publication at the end of the project. Guidelines on implementing data management planning for postgraduate researchers in the context of a DTC will be of great value to academics currently applying for DTC funding from EPSRC. Finally, guidance on the development of a data management policy, in which the experiences of the project team are documented, is also in development and will be published when the University's data management policy receives final approval.

<u>Improved regional and national collaborations</u>

Early in the project, Research360 established links with other Jisc-funded MRD projects in the region: 'data.bris' at the University of Bristol; 'Open Exeter' at the University of Exeter; and 'Managing Research Data: a pilot study in Health and Life Sciences' at the University of the West of England. Strong links were established through regular contact either at regional meetings hosted at each institution, or at frequent Jisc MRD programme events. Through these links, experiences and ideas were shared. The project coincided with the formation of a regional collaboration (GW4) between the universities of Bath, Bristol, Cardiff and Exeter. As a result of improved data management activities at all four universities, a data service working group has established to formalise and strengthen the links created during the Jisc-funded MRD projects, thus ensuring that the immediate impact of the project will continue to have long-term regional benefits.

2.5 Future Impact

<u>Compliance with EPSRC Expectations</u>

The Research360 project has identified a number of key ways in which it will continue to impact data management within the University of Bath. Most importantly, by securing permanent resource to continue with two dedicated full-time data management roles, implementation of the project's *Roadmap for EPSRC* will be enabled, ensuring compliance with funder requirements. Tracking the impact of the Roadmap's implementation will be achieved via reporting progress against milestones to the new Research Data Sustainability Group.

<u>Pilot Data Repository Development</u>

A critical aspect of the Roadmap's implementation will be to build on a new pilot institutional data repository, initiated and piloted by the Research360 project. Continued development of this pilot repository will enable researchers to archive, and where possible publish, their research data, or to register the location of either non-digital data or data deposited elsewhere. The pilot data repository will enable monitoring of access to and downloads of published research data, in order to track impact. In addition, longer term plans to register with DataCite will allow the repository to mint Digital Object Identifiers; this will assist in the publication of recommended formats for data citation and will enable core DataCite metadata fields to be harvested by developing national and international data catalogues. This will, in turn, promote data discovery and re-use and ultimately will enable data citations to be tracked.

Selection of EPrints for the pilot data repository would facilitate integration with the existing research infrastructure, particularly the EPrints based publications repository that has already been integrated with the University's CRIS, should this option be desired. Such an integration would allow links to be made between project funding, published articles, equipment and research data. Such links would allow the CRIS to be queried for metrics such as the number of research projects that have published data and the proportion of published articles with supporting data available. It would also enable compliance with open access or individual funder policies to be demonstrated. Importantly, integration with the CRIS would allow many metadata fields to be prepopulated during ingest of new datasets, thus streamlining the deposit workflow for researchers and reducing barriers to data publication.

<u>Advocacy and Guidance</u>

Advocacy and training materials developed by the Research360 project provided a general overview of data management issues, with links to external sources of additional information frequently used. As such, the resources were designed to be easy to maintain and remain relevant once the project had finished. It is anticipated that these resources will not only continue to be used, but will be further developed to provide additional, more detailed information, some of which is already being requested by researchers. The impact and suitability of these resources will be monitored in a number of ways. For RDM training workshops, feedback is solicited from participants after the event and this

mechanism will be used both to monitor changes in participants' behaviour and to inform the development of new resources. As the new RDM website is expanded, metrics such as the number of unique page views, time spent on each page, bounce rate and exit rate for each page will be used to monitor how researchers are moving through and using the resources, to enable the range of information provided and its structure to be improved.

3 Conclusions

General Conclusions

A number of conclusions can be drawn from the experiences of the Research360 project. In general, the project team found that a considerable amount of time is required to explore data management issues, find solutions, and develop resources, whilst at the same time building and maintaining a supporting network of both internal and external contacts. As such, it would not have been possible to deliver the main outcomes of the project had external resource not been provided to fund dedicated roles. Other core responsibilities meant that, in general, staff had little free resource to commit to new data management tasks, a problem that was exacerbated by activities such as the REF 2014, Open Access requirements and staff restructuring.

The project concluded that the approach taken to policy development could have been improved. Firstly, as discussed under Section 3.2, the collaborative approach taken to policy development resulted in excessive re-writing of some aspects of the policy, whilst the overall aim of the policy was lost sight of. On reflection, early clarification of the policy's high-level requirements and the importance of consultation with researchers are two factors that would have considerably helped the development and approval process. The former would have prevented circular revisions on the extent to which open data publication should be mandated. The latter would have identified the optimal format, tone, scope and guidance earlier in the approval process. The project also concluded that it was worth investing more time in policy development, to ensure that the complex issues relating to management of collaborative research data were correctly handled, rather than approving an aspirational set of principles that could not be formalised as a University policy, and then having to repeat the process in the near future.

The project also concluded that the drivers of data management infrastructure development may depend on the part of the data lifecycle being supported and include an element of iterative feedback. For example, as discussed in Section 3.3 an effective driver of repository development was demand from researchers for a platform to enable data publication. In contrast, to encourage researchers to improve sub-optimal data storage and data sharing practice it is necessary to ensure a supporting infrastructure is already in place - it is unreasonable to tell researchers not to manage their data in particular way without providing them with an alternative solution. For data storage, the project found that uptake of existing reliable, secure centrally managed storage could be enhanced by providing and promoting at least some of this storage 'free' to researchers. As demand for this storage continues to increase beyond current capacity, it is anticipated that this demand will drive further expansion in storage infrastructure. Similarly, the project discovered a great demand amongst researchers for shared research space to enable collaboration with external partners. However, in order to tempt researchers away from cloud services such as Dropbox, an institutional alternative must be available. To address concerns relating to cloud services and data covered by the Data Protection Act or confidential collaboration agreements, this service would have to be secure and work with local infrastructure.

Conclusions Relevant to the Wider Community

Of relevance to the open source community, was the project's conclusion that the Sakai VRE documentation is so poor that it is proving to be a barrier to further development work. Such work would be unfeasible due to the additional costs incurred, either by the extended time needed for

developers new to Sakai to become familiar enough with the environment to commence work, or by the higher salaries likely to be commanded by a declining number of experienced Sakai developers.

Funding councils and other HEIs should note the project's conclusion that the information considered of interest in a DMP to both researchers and the university differs from that frequently requested in DMPs requested by funders at the grant application stage. For example, funders are understandably interested in ensuring that existing data are not being unnecessarily recreated. However, researchers tend to see requests to demonstrate the requirement for new data collection as unnecessary, or even insulting, as they themselves have a clear overview of current progress in their fields. Similarly, whilst funder DMPs do not tend to require quantification of storage requirements, this information is of value for institutional annual planning and should also be used to justify any costs requested to support management and storage of data during a project. Further, much of the information required by funder DMPs remains the same between projects and might therefore be better suited to an institutional, departmental or group level data management policy. Although it should be possible to re-use or copy such content between DMPs, this does not represent an efficient use of researchers' time. Currently, the role of the DMP appears primarily to promote data publication, thus ensuring that data arising from funder investment is made widely available. Once the culture of data sharing has changed, it may be necessary to review the role of the DMP.

Related to this was how the diversity of DMP templates and data policies mandated by the UK funding councils impacted on institutional support. In order to assist all researchers in writing DMPs, a range of funder-specific guidance and tools are needed, requiring resource to both provide and maintain. The project team was uncertain as to whether this investment of resource is worthwhile. For example, how these DMPs are then used during the grant application process has been a subject of discussion both within the project team, at Jisc MRD programme workshops and at the recent RDMF 'A conversation with the funders' workshop. The project determined that, in general, DMPs are not always used as filters when allocating research grant funding, although NERC, BBSRC and MRC are notable exceptions. As such, the project concluded that investment of considerable resource in data management planning could not currently be justified by the potential for an increase in grant funding.

Further, the project team felt it was unclear whether the peer review panels could be relied upon to have the necessary data management expertise required to accurately evaluate the suitability of DMPs they assessed. The project team considered a number of ways by which this could be improved. One approach would be to separate review of DMPs from the peer review of funding applications, so that this was instead carried out by specialist technical staff employed by the funding bodies. Alternatively, funding bodies asking researchers to review DMPs have a responsibility to provide training to support this. Institutions could contribute in two ways: Firstly, institutions, like funding bodies, could provide specific DMP-review training for those senior academics who frequently sit on grant review panels. Secondly, the need for this DMP-specific review training should reduce over time, as a result of ongoing advocacy and training activities within institutions, resulting in improved awareness of DMP requirements and a general cultural change in data management practice. However, although an institution would benefit from the latter, it is not clear how this investment in advocacy and training could be directly linked to increased revenue from successful grant applications. Unless other institutions undertook a similar approach, any improved DMPs submitted could potentially still be reviewed by untrained academics.

Conclusions Relevant to Jisc

Of relevance to the Jisc-funded DCC was the project's conclusion that the current DMPonline tool is unsuitable for use by DTC-based or other postgraduate research students. The project ran a focus group for DTC-based PGR students, where DMPonline was evaluated alongside three other DMP templates: DataTrain's postgraduate DMP form; an early version of David Shotton's 'Twenty Questions' template; and the Research360 project's Data Management Plan for Postgraduate Researchers. The project concluded that the amount of detail requested by DMPonline was off-putting

to PGR students, who felt that very little of it was relevant to their projects. Further, the students frequently did not understand what was being asked.

Similarly, from a series of DMP workshops, targeted at different funding councils, the project concluded that the current DMPonline tool could not be recommended for use creating data sharing plans or technical appendices mandated for grant applications. This was particularly so where a funder had provided a specific DMP template, or where a page limit restricted the length of the DMP. It is important to note however, that the DCC are already addressing these concerns and have planned a redesign of the DMPonline tool so that it more closely aligns with funder requirements.

4   Recommendations

The Research360 project makes the following recommendations:

1.   Whilst the project has successfully launched a pilot data service that provides for general data management advice, it is becoming increasingly apparent that there is still a considerable amount of work required to fully support all aspects of research data management. In particular, more detailed discipline- and role-specific guidance and training are required. This should particularly cover selection and retention of data; licencing issues; data repository specifications; and the specific requirements of data repository policies, end user agreements and terms of use. The project therefore recommends that future Jisc MRD programmes are funded, to build on the outputs of the current programme and to develop transferable resources covering these issues. The project acknowledges that the DCC or DataCite at the British Library might represent ideal forums for bringing this type of information together.

2.   As acknowledged in Section 1, the Research360 team found that the support of the Jisc MRD programme and wider RDM community was a valuable asset during the project, providing ideas, encouragement and solutions. The project therefore recommends that Jisc continue to maintain both the JISCMRD@JISCMAIL.AC.UK and RESEARCH-DATAMAN@JISCMAIL.AC.UK mailing lists as a centre for communication for the growing UK and international data management communities respectively.

3.   The DCC are already planning a redesign of the DMPonline tool, which will address the conclusions drawn in Section 4. Once this has been completed, the Research360 project recommends that funding councils should update their data policies so that DMPs generated using the DMPonline tool are accepted either in place of, or alongside, templates provided by the funding councils. Acceptance of DMPs created using the DMPonline tool should be clearly stated in funder polices and guidance. The project recognises the need for co-ordination between the funders and the DCC for this recommendation to be achieved.

4.   The project concluded that inconsistency between funder policies presented institutional RDM staff with difficulties in providing guidance and tools to meet these policies, creating a significant barrier to uptake. In order to make the most efficient use of limited institutional resources, the project therefore recommends that UK funding bodies align their policies and requirements for data management. In some cases, this may be as simple as using consistent terminology. For example, data management plans are also called 'technical plans', 'data sharing statements' or 'data sharing and preservation strategies'. Whilst these terms might be understood by those experienced in data management, they can often be confusing to researchers who are completely new to the process.

5.   As discussed in Section 4, to improve the process of DMP peer review the project recommends that, if they do not already do so, funding bodies should provide specific guidance and training to support researchers when peer reviewing DMPs. Associated with this, the project recommends that Jisc continue to support the development of data management infrastructure, including advocacy and training, particularly at institutions where this has not yet been started. This will

ensure that eventually *all* researchers from *all* institutions will not only possess the skills necessary to manage their own data, but they will also be able to evaluate the appropriateness of other researchers' plans for data management.

6.   Based on the project's experience of re-using existing materials – successfully in the case of the *Roadmap for EPSRC*, but less so for the online training module - the project recommends that future projects and other universities undertaking similar data management initiatives should directly utilise outputs from Jisc MRD projects as much as possible. As demonstrated in Section 3.3, substantial costs savings could be made using this approach.

7.   Based on the project's experience of developing for the Sakai VRE, discussed in Section 3.3, the project recommends that open source projects such as Sakai should enhance their documentation, particularly guidance aimed at future developers. Of particular relevance to Jisc is the project's acknowledgement that it may be necessary for funding bodies to provide support for such activities. Please note that additional recommendations relevant both to future Sakai developers, and to the wider Sakai community, have been published in the *Sakai-SWORD2 Integration Development Report*. These recommendations are included as Appendices 2 and 3 in Section 8 of this project report.

8.   As discussed in Section 3.5, the project's pilot institutional data repository might eventually be integrated with the University's existing research infrastructure, notably the CRIS. It is known from attendance at Jisc MRD programme workshops that other institutions have similar plans, both to enable links to be made between research inputs and outputs, and to facilitate data deposit by researchers. However, both CRIS and data repositories are generally in the early stages of development, with a number of commercial CRIS systems in use and institutions employing a range of bespoke solutions to enable data archive and publication. As such, integration is likely to be complex, possibly including a number of intermediate steps. For example, metadata schema would have to be mapped between the two systems and, should the CRIS be used to directly handle data as well as metadata, it would probably have to be customised both to enable ingest and download of large datasets and to handle the substantial underlying storage requirements. Unless a CRIS immediately superseded a data repository, it would be necessary to determine how the two systems might work together – either one system could be used for the ingest of data and metadata, with both pushed or pulled to the other system for publication and external access, or alternatively one system might handle ingest and publication of data and metadata, with the other acting as a metadata catalogue but not handling the full data records. The project therefore recommends that over the next few years the interface between the RIM and RDM communities, including both users and software providers, are explored and supported by Jisc, the DCC and the wider community.

9.   As mentioned in Sections 2 and 4, the project found that researchers identified sharing of research active data as a particular barrier to good data management practice. Based on the project's conclusion that a secure, institutional alternative to Dropbox must be available, it is therefore recommended that Jisc support rapid development of such a solution. Alternatively, if a solution already exists, either open source or commercial, Jisc should highlight this to the wider community and provide support for its implementation.

10. As discussed in Section 3.3, the Research360 project found that researchers preferred didactic over discursive training workshops. The project therefore recommends that other Universities developing data management training materials for researchers use a similar didactic approach combined with structured exercises. The project also recommends that training materials include practical advice, examples, links to further information and recommended software and tools to support all aspects of data management.

5   Implications for the future

Once the Research360 project has finished, it will be possible to contact members of the project team via research-data@bath.ac.uk. Published project reports will be archived in Opus, the University of Bath's online publications store, and outputs completed before June 2013 will also be available from the archived UKOLN Research360 project page, at http://www.ukoln.ac.uk/projects/research360/. For both the terminated Sakai development work and some initial work to develop an add-on for EPrints that enables interaction with a Hitachi Content Platform filestore, code and supporting documentation has been made available primarily via Bitbucket, at https://bitbucket.org/research360, which offers its paid-for plan free to academic users and enables views to be tracked by google analytics.

Other project outputs, such as the University's RDM website (available at go.bath.ac.uk/research-data), the new online training module and workshop and the pilot institutional data repository, will form the basis of the new data management service. These project outputs will therefore be maintained, developed, expanded and used by the permanent data management roles established by the project. The primary implication of this is that the University of Bath will have an on-going data management service to support researchers in the management of their research data.

It will be possible for other professionals in the field to continue development of some project outputs. For example, Xerte was selected to implement the new online training module not only because it is open source, enabling the module to be freely reused by other institutions, but also because the same technology was used to develop the MANTRA training module. This means that future re-users could combine content from both courses. Although the Sakai development work was unsuccessful, the valuable lessons learnt have been documented and the code made available, as other developers may wish to attempt similar work in the future.

6   References

Beagrie, N., Beagrie, R., Rowlands, I., 2009. Research Data Preservation and Access: The Views of Researchers. *Ariadne* [online], 60. Available from: http://www.ariadne.ac.uk/issue60/beagrie-et-al/ [Accessed 4 July 2013].

Beitz, A., Dharmawardena, K., Searle, S., 2012. *Research Data Management Strategy and Strategic Plan 2012–2015* [online]. Victoria: Monash University. Available from: http://monash.edu/library/researchdata/about/strategy/ [Accessed 8 July 2013]

DataTrain Project, 2011. *Data Management Plan for Post-Graduate Research Projects* [online]. Cambridge: University of Cambridge. Available from: http://www.lib.cam.ac.uk/preservation/datatrain/Post-Graduate_DMP_Form.pdf [Accessed 8 July 2013]

EPSRC, 2011. *EPSRC Policy Framework on Research Data* [online]. Swindon: EPSRC. Available from: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx [Accessed 8 July 2013].

Guy, M., 2013. *Increasing Participation in Internal RDM Training Sessions. DCC Briefing Papers.* DCC RDM Services Case Studies. Edinburgh: Digital Curation Centre. Available from: http://www.dcc.ac.uk/resources/developing-rdm-services [Accessed 8 July 2013].

Jones, K., 2011. *Assessing institutional data storage and management using the Data Asset Framework (DAF) methodology at the University of Bath.* Bath: University of Bath. Available from: http://opus.bath.ac.uk/24960/ [Accessed 8 July 2013].

Open Exeter Project Team, 2012. *Summary Findings of the Open Exeter Data Asset Framework (DAF) Survey.* Exeter. University of Exeter. Available from: http://hdl.handle.net/10036/3689 [Accessed 8 July 2013]

Pink, C., Cope, J., Jordan, K., Jones, K., (2013). *Faculty-Industry Data Requirements Report.* Unpublished.

Shotton, D., 2012. *Twenty Questions for Research Data Management* [online]. Oxford: University of Oxford. Available from: http://imageweb.zoo.ox.ac.uk/pub/2012/publications/Shotton-Twenty_Questions_for_Research_Data_Management.docx [Accessed 8 July 2013].

University of Bath, 2009. *Charter of Incorporation* [online]. Bath: University of Bath. Available from: http://www.bath.ac.uk/about/pdfs/CharterAug09.pdf [Accessed 8 July 2013].

7   Appendices (optional)

Appendix 1: Glossary of Acronyms

BUCS	Bath University Computing Services
CRIS	Current Research Information System
DAF	Data Asset Framework
DCC	Digital Curation Centre
DMP	Data Management Plan
DTC	Doctoral Training Centre
EPSRC	Engineering and Physical Sciences Research Council
KRDS	Keeping Research Data Safe
PGR	Postgraduate Research student
RDM	Research Data Management
RDMF	Research Data Management Forum
RDSO	Research Development and Support Office
REF	Research Excellence Framework
RIM	Research Information Management
VLE	Virtual Learning Environment e.g. Moodle
VRE	Virtual Research Environment e.g. Sakai

Appendix 2: Recommendations for Improvements to Sakai Documentation

This appendix contains recommendations for improvements to Sakai documentation, which are due to be published as an appendix to the *Sakai-SWORD2 Integration Development Report*, and have also been passed separately to the Sakai community.

Background

The Jisc-funded Research360 project at the University of Bath aimed to integrate data deposit seamlessly into the research workflow, by enabling data to be transferred from a virtual research environment directly into a data repository. The first stage of this involved developing an integration between the Sakai virtual research environment and the SWORD2 deposit specification. Development work on the integration between SWORD2 and Sakai proved to be technically far more challenging than anticipated. As such, the work was suspended in late 2012. A frequently encountered difficulty during the work was the poor quality of Sakai's documentation. In order to contribute to improvements in this area, this document contains a list of recommendations that, if implemented, are likely to improve the experience of future developers who are new to Sakai.

Recommendations

The following list suggests some ways in which improvements could be made, ordered by how easy it is anticipated they would be to implement.

1. On the Sakai wiki, replace the search box provided by the Confluence software with a domain-restricted search box for a major search engine e.g. Google – for example, a Confluence search for "authorization" returns many results for "authoring" instead. Ensure that the Sakai wiki pages are indexed by Google and other major search engines.

2. Update documentation for the *sakai.properties* file used for localisation of a Sakai installation. There are currently two sources of information: example files (one containing commonly needed

customisation, the other every possible item that can be customised), and an out of date Word document listing the customisable values with minimal information on what they mean. Not all fields have clear functions or acceptable values that can be inferred from their names. This documentation on the Sakai wiki is quite out of date.

3. Volunteers from the community should regularly scan the Sakai wiki to identify and then either fix or remove broken links.

4. Documentation matching frequently asked or anticipated questions, such as how the Sakai authorisation model works, should exist in the Sakai wiki.

5. Easy to find and regularly managed pathways into the Sakai wiki should be created and these should be linked directly from the Sakai website page for developers. These could pick up common issues for developers. Although the Programmer's Café goes some way toward this, it is not yet sufficient alone. In particular, links to the Programmers Café and other developer areas of the Sakai wiki are conspicuously absent from the Getting Started page for Technical Contributors[7].

6. Discoverability of Javadocs should be improved, particularly for classes which are shipped as binary overlays and are therefore not in the standard source distribution. This might involve a central index of all versions of each module, along with some way of determining which module a particular class belongs to. Care should be taken over the page titles of generated Javadocs to make them more easily discoverable and identifiable through major search engines e.g. Google. The current alternative necessitates browsing through multiple locations found at the Programmer's Cafe page on Javadocs[8]. Without knowledge of the project it is not immediately obvious whether a particular class will be part of the API, the kernel, or needs to be found in the individual release Javadocs. It is acknowledged that this may require a change to the way that Javadocs are generated.

7. Developers should be encouraged to include more comments in their work and to add descriptions to functions, constants, etc. If correctly formatted, the latter would then automatically appear in generated Javadocs.

8. The Sakai wiki would benefit from some usability testing of access to information.

9. The Sakai wiki would benefit from both removal of out-of-date information and the addition of more up-to-date information. In addition, links from discussion documents to information about the resolution of the discussion would be particularly useful.

10. Some overarching management of the Sakai wiki content and structure might help to reduce proliferation of out-of-date pages.

11. The wiki pages on development environment set up should be altered to make clear that using an integrated development environment (IDE) such as Eclipse is optional, and not required for successful Sakai development. The high memory requirements of IDEs are particularly exposed by Sakai, which itself requires a lot of memory.

Appendix 3: Recommendations to future Sakai developers

This appendix contains general recommendations for the Sakai community, and specific recommendations aimed at future developers intending to work with Sakai. The recommendations are due to be published in the *Sakai-SWORD2 Integration Development Report*.

Two general recommendations are made to the wider Sakai development community working in this area:

---

[7] http://www.sakaiproject.org/technical-contributors
[8] https://confluence.sakaiproject.org/display/BOOT/Javadocs+for+the+Sakai+API

1. Widely used open source projects such as Sakai should enhance their documentation, particularly guidance aimed at potential developers. It may be necessary for funding bodies to provide support for such activities. A full list of recommended improvements to Sakai documentation is provided in Appendix 1.

2. Sakai should ensure that all relevant files required to create a Sakai instance, and to develop new code, are included in every relevant branch of the subversion code repository. In particular, the *content-api* library must be included in future releases.

The following recommendations are aimed at developers who plan to work on Sakai:

1. If there is freedom to choose which version of Sakai is used, consider taking advantage of the widget architecture for Sakai OAE where possible rather than working with the more complex API in Sakai CLE.

2. Install a dedicated development copy of Sakai rather than relying on an existing production server.

3. Rather than carrying out the compilation to set up a demo or test site, instead consider a site installed from a pre-compiled binary. When testing modified code in the site, replace the modules that have been amended with compiled versions from the updated code. Configure Java, Tomcat and Maven to use as much memory as is available when compiling and when the Tomcat server is running: This must be at least 2GB of RAM, although 4GB is preferable.

4. If a fully compiled version of Sakai is required, then it should be installed using components downloaded directly from the relevant sources or software development kits (SDK) e.g. Tomcat and the Oracle/Sun Java SDK, rather than using packaged versions of these or other SDKs/Java VMs. If errors occur it can be problematic to re-install over an existing installation. Instead, a clean Tomcat needs to be set up for Sakai from scratch each time as a non-standard configuration of Tomcat is needed for Sakai to run. This can be achieved with a single installation of Tomcat using multiple instances[9].

5. When developing for the Sakai CLE, it is advisable not to install Sakai from source, nor to attempt to add the whole of Sakai to the Eclipse integrated development environment (IDE). When compiling for import into Eclipse, use the *sakai:deploy* profile and omit those Sakai modules that are not relevant to the development task from both the compilation and import.

6. The requirement for high memory usage of integrated development environments (IDEs) is particularly exposed by Sakai. If Eclipse is used, the source code must be compiled to import into the IDE. For this purpose, the development machine will require more than 3GB RAM. It is therefore worth considering not using an IDE at all. Development in a standard text editor, combined with standard Unix command line tools, was found to be easier than using Eclipse.

7. A relatively easy way to identify the Java code used to produce a specific page or function is to analyse the HTML source code displayed when a task is carried out on a Sakai server using the web browser. Use of the Unix command line tools *grep -r* and *find* was found to be faster than the search tool in Eclipse for finding relevant code in the Sakai source, the latter tool having a tendency to hang and not complete searches.

8. When working with the CLE, remember that it is a complex system including a lot of legacy code and allow additional time to learn how it works, looking in detail at the classes in the module(s) to be modified. As class files can be very large and complex, this is likely to be a time consuming process.

9. Join the Sakai development mailing list and make use of it earlier rather than later when

---

[9] See the "Advanced Configuration - Multiple Tomcat Instances" in RUNNING.txt in the root of a standard Tomcat installation.

problems arise. Contribute to improvements in Sakai documentation by updating key information with the helpful advice provided by the Sakai community.