

Project Identifier: #rdtk\_herts - Service Oriented Toolkit for Research Data Management

Final Report, Version: 0.5, June 2014

Contact: Dr Bill Worthington, [w.j.worthington@herts.ac.uk](mailto:w.j.worthington@herts.ac.uk)

## JISCMRD Final Report - Service Oriented Toolkit for Research Data Management

Project Information			
Project Identifier	<a href="#">RDTK</a>		
Project Title	Service Oriented Toolkit for Research Data Management		
Project Hashtag	#rdtk_herts		
Start Date	03/10/2011	End Date	30/6/2013
Lead Institution	University of Hertfordshire		
Project Director	Dr David Ford, Chief Technology Officer, University of Hertfordshire		
Project Manager	Dr William Worthington, Group Leader, Enterprise Applications Consultants, University of Hertfordshire (known as Bill)		
Contact email	<a href="mailto:w.j.worthington@herts.ac.uk">w.j.worthington@herts.ac.uk</a>		
Partner Institutions	Digital Curation Centre		
Project Webpage URL	<a href="http://www.herts.ac.uk/research-data-toolkit">http://www.herts.ac.uk/research-data-toolkit</a>		
Programme Name	Managing Research Data		
Programme Manager	Dr Simon Hodson		

Document Information			
Author(s)	Bill Worthington		
Project Role(s)	Project Manager		
Date	05/06/2014	Filename	rdtk-final-report-v05.pdf
URL	<a href="http://hdl.handle.net/2299/13636">http://hdl.handle.net/2299/13636</a>		
Access	Public		

Document History		
Version	Date	Comments
0.1	14/06/2013	Draft for final RDTK Steering Group
0.2	21/06/2013	Revised Draft for submission to JISC. New: Project Summary, Section 11 Engagement, appendices. Revised: list of outputs, learning in section 3.2.5 and 3.2.9
0.3	09/07/2013	Added idealised cloud storage diagram
0.4	20/05/2014	Final version for repository deposit. Updated links in Table of Outputs, and links to current services throughout.
0.5	05/06/2104	Updated with persistent identifier <a href="http://hdl.handle.net/2299/13636">http://hdl.handle.net/2299/13636</a>

Note: to facilitate the flow of the narrative of this report we refer to University of Hertfordshire as UH, often and interchangeably.

## Table of Contents

<b>1</b>	<b>ACKNOWLEDGEMENTS</b>	<b>3</b>
<b>2</b>	<b>PROJECT SUMMARY</b>	<b>3</b>
<b>3</b>	<b>AN ACCOUNT OF THE PROJECT</b>	<b>4</b>
3.1	PROJECT OUTPUTS AND OUTCOMES	4
3.2	ACTIVITY AND LEARNING	6
3.2.1	<i>WP1: Audit current UH good practice data management</i>	6
3.2.2	<i>WP2: Cloud storage pilot (Health sciences)</i>	7
3.2.3	<i>WP3: Document Management pilot (Health sciences)</i>	10
3.2.4	<i>WP4: Thirty party repository pilot (Physics)</i>	11
3.2.5	<i>WP5: Applicability for the Humanities (History)</i>	12
3.2.6	<i>WP6: Review data protection and IPR issues</i>	13
3.2.7	<i>WP7: Review long term sharing, storage and access issues</i>	13
3.2.8	<i>WP8: Data Repository (internal/external partners)</i>	13
3.2.9	<i>WP9 and WP10: The Research Data Toolkit</i>	14
3.2.10	<i>WP11A: Programme Engagement (external)</i>	15
3.2.11	<i>WP11B Programme Engagement (internal)</i>	16
3.3	IMMEDIATE IMPACT	16
3.4	FUTURE IMPACT	17
<b>4</b>	<b>CONCLUSIONS</b>	<b>18</b>
<b>5</b>	<b>RECOMMENDATIONS</b>	<b>18</b>
<b>6</b>	<b>IMPLICATIONS FOR THE FUTURE</b>	<b>19</b>
<b>7</b>	<b>REFERENCES</b>	<b>20</b>
	<b>APPENDICES</b>	<b>20</b>

## 1 Acknowledgements

The Service Oriented Toolkit for Research Data Management project was co-funded by the JISC Managing Research Data Programme 2011-2013 and The University of Hertfordshire. We owe thanks to all the participants of JISCMRD who shared their experience in a truly collaborative effort. The project benefited in particular from the work of the Digital Curation Centre and the Universities of Essex, Leeds, Leicester, Lincoln, Nottingham, and Southampton. Within the University of Hertfordshire we should acknowledge the role of Prof Owen Davies, Prof Janet Drew, Dr Karin Friedli, Prof Martin Hardcastle, and Dr David Wellsted for their advocacy and interface to our research community. Thanks are also due to the senior management of Information Hertfordshire for providing the scope in which the project could explore and act.

## 2 Project Summary

The University of Hertfordshire (UH) conducts world leading research across several disciplines, which generates substantial amounts of data of many different types and scale, and significant project documentation. The University was aware of the data management and sharing policies being developed by a variety of research funders, and had begun to produce Research Data Management (RDM) specific policy and support prior to the advent of JISCMRD2.

This project focused on the realisation of practical benefits for operationalising an institutional approach to good practice in RDM with strong transferable value.

The objectives of the project were to audit current best practice, develop technology demonstrators with the assistance of leading UH research groups, and then reflect these developments back into the wider internal and external research community via a toolkit of services and guidance. The overall aim was to contribute to the efficacy and quality of research data plans, and establish and cement good data management practice in line with local and national policy.

We were largely successful in these objectives despite encountering a less mature technology market than expected and an equally young discipline, which often revealed more questions than answers. It is too early to know whether the aims were similarly successful, since the derived benefit will take some years to accrue.

We 'operationalised' less new services than we might have hoped, but achieved clarity about what we must do and how to move forward. This clarity comes from funders, from our research community and from ourselves as service providers. As a result the project has been, and will be, an effective agent of change at several levels: in intervening with day to day practice via advice and solutions; in understanding and developing the university's facilities, particularly in the context of a hybrid-cloud infrastructure; and in defining and lobbying on issues in a national context.

The project has participated fully in the JISCMRD event programme, and benefited from and contributed to it in equal measure. The shared experience and audit results across JISCMRD show close commonality, so we believe the learning delivered in this report, which we think is considerable, will be applicable and of use across the sector.

### 3 An account of the project

The activity on the project was organised via a set of work packages, often overlapping or informing each other, and evolving to meet the needs or themes that were revealed as the project progressed. In this section we will first list the outputs and then describe the activity and learning in each work package.

#### 3.1 Project Outputs and Outcomes

Blog	WP1	Survey based on Digital Asset Framework	<a href="http://bit.ly/18QUZR9">http://bit.ly/18QUZR9</a>
Survey	WP1	Survey results	<a href="http://bit.ly/1ao74vy">http://bit.ly/1ao74vy</a>
Report	WP1	Survey analysis	<a href="http://bit.ly/128uGMK">http://bit.ly/128uGMK</a>
Blog	WP1	UH Research Data Policy in a nutshell	<a href="http://bit.ly/14cXC9w">http://bit.ly/14cXC9w</a>
Artefact	WP1	Interview protocol, used by project analyst and RDM champions	<a href="http://bit.ly/12Jr9KZ">http://bit.ly/12Jr9KZ</a>
Case studies	WP1	12 Case Studies	<a href="http://bit.ly/19MjnD3">http://bit.ly/19MjnD3</a>
Review	WP2	Review of cloud storage services: features, costs, issues for HE	<a href="http://bit.ly/12Jn2yz">http://bit.ly/12Jn2yz</a>
Blog	WP2	Files in the cloud	<a href="http://bit.ly/R583lf">http://bit.ly/R583lf</a>
Test data	WP2	Files transfer rate tests	<a href="http://bit.ly/1266WsJ">http://bit.ly/1266WsJ</a>
Blog	WP2	Analysis of barriers to use of local networked storage	<a href="http://bit.ly/12Glegg">http://bit.ly/12Glegg</a>
Blog	WP2	Hybrid-Cloud model: when the cloud works and the attraction of Dropbox et al.	<a href="http://bit.ly/Xvmidr">http://bit.ly/Xvmidr</a>
Blog	WP2	Hybrid-Cloud example: Zendto on Rackspace, integrated with local systems	<a href="http://bit.ly/11In83q">http://bit.ly/11In83q</a>
Service	WP2	UH file exchange	<a href="https://www.exchangefile.herts.ac.uk/">https://www.exchangefile.herts.ac.uk/</a>
Blog	WP2	Cost of ad-hoc storage	<a href="http://bit.ly/19ilycQ">http://bit.ly/19ilycQ</a>
Blog	WP2	Cost of data loss event	<a href="http://bit.ly/13RSckb">http://bit.ly/13RSckb</a>
Blog	WP2	Reflection on use of Rackspace CloudFiles	
Blog	WP2	Data Encryption	<a href="http://bit.ly/XxDoEM">http://bit.ly/XxDoEM</a>
Training		Data Encryption workshop	<a href="http://bit.ly/11rwLXA">http://bit.ly/11rwLXA</a>
Training	WP2	Data Encryption guide	<a href="http://bit.ly/QHyN2y">http://bit.ly/QHyN2y</a>
Blog	WP3	Document Management for Clinical Trials	<a href="http://bit.ly/15cfT5K">http://bit.ly/15cfT5K</a>
Artefact	WP3	eTMF - electronic Trial Master File, 1954 legacy documents scanned	no public access
Artifact	WP3	Research Project File Plan	<a href="http://bit.ly/11InVkW">http://bit.ly/11InVkW</a>
Workflow	WP3	Post award storage allocation	

Workflow	WP2, WP3	Request 'Research Storage' Form	<a href="http://bit.ly/17V7J8t">http://bit.ly/17V7J8t</a>
Workflow	WP2, WP4	Research Grant and Storage Process	<a href="http://bit.ly/14kvCB0">http://bit.ly/14kvCB0</a>
Workflow	WP2, WP4	Request 'Research Storage' Workflow	<a href="http://bit.ly/12d2aJP">http://bit.ly/12d2aJP</a>
Service	WP2, WP3	R: (R drive), workgroup space with external access	access by workgroups
Service	WP3	DMS, workgroup space with external access	access by workgroups
Dataset	WP5	4 Oral history datasets, ~300 interviews, 125GB	<a href="http://bit.ly/uh-hhub">http://bit.ly/uh-hhub</a>
Dataset	WP5	1 Leisure studies dataset, SPSS survey, interviews, transcripts, 8GB	in preparation
Blog	WP6	Comparison of data licenses	<a href="http://bit.ly/12DmXfR">http://bit.ly/12DmXfR</a>
Report	WP6	Comparison of data licenses	<a href="http://bit.ly/13NC7gA">http://bit.ly/13NC7gA</a>
Service	WP8	UHRA repository improvements phase 1	<a href="http://uhra.herts.ac.uk/">http://uhra.herts.ac.uk/</a>
Blog	WP8	DOIs for datasets, includes mind map	<a href="http://bit.ly/QonFoN">http://bit.ly/QonFoN</a>
Workflow	WP8	Deposit/access criteria for data with a levels of openness	<a href="http://bit.ly/12cUjrq">http://bit.ly/12cUjrq</a>
Service	WP9	RDM micro site (aka Research Data Toolkit), 100+ pages and pdfs of RDM guidance	<a href="http://bit.ly/uh-rdm">http://bit.ly/uh-rdm</a> <a href="http://www.herts.ac.uk/rdm/">http://www.herts.ac.uk/rdm/</a>
Report	WP11A	Register of Programme engagement at external events, estimated audience 480, ~300 individuals	Appendix A
Blog	WP11A	Programme engagement: 38 Blog posts	<a href="http://research-data-toolkit.herts.ac.uk/">http://research-data-toolkit.herts.ac.uk/</a>
Presentation	WP11A	Programme engagement: 14 Presentations	
Presentation		Association of Research Managers and Administrators Conference 2013	<a href="http://bit.ly/ZXv8RK">http://bit.ly/ZXv8RK</a>
Presentation		UH RDM Stakeholder briefing June 2012	<a href="http://bit.ly/11KkJGo">http://bit.ly/11KkJGo</a>
Presentation		UH Health and Human Sciences research forum July 2012	<a href="http://bit.ly/15cDUKb">http://bit.ly/15cDUKb</a>
Presentation		JISCMRD progress workshop Nottingham 2012: storage	<a href="http://bit.ly/10qpry3">http://bit.ly/10qpry3</a>
Presentation		JISCMRD progress workshop Nottingham 2012: repository	<a href="http://bit.ly/126zjab">http://bit.ly/126zjab</a>
Presentation		JISCMRD progress workshop Nottingham 2012: training	<a href="http://bit.ly/15cH1lj">http://bit.ly/15cH1lj</a>

		JANET/JISCMRD Storage Requirements workshop Paddington 2013	<a href="http://bit.ly/12QFu9S">http://bit.ly/12QFu9S</a>
Presentation		JISCMRD benefits evidence workshop Bristol 2013	<a href="http://bit.ly/ZXE09Y">http://bit.ly/ZXE09Y</a>
Presentation		JISCMRD progress workshop Aston 2013: training	<a href="http://bit.ly/11t3Lg0">http://bit.ly/11t3Lg0</a>
Presentation		JISCMRD progress workshop Aston 2013: agent of change	<a href="http://bit.ly/13NVlgH">http://bit.ly/13NVlgH</a>
Presentation		JISCMRD progress workshop Aston 2013: storage	<a href="http://bit.ly/19Juixf">http://bit.ly/19Juixf</a>
Report	WP11B	Register of programme engagement at UH events: interviews (~60), meetings, seminars, workshops. Total attendance 400, est 200-300 individuals	Appendix B
DMP	WP11B	10 data management plans, facilitated by RDM champions and Research Grants Advisor	limited public access
Report	WP12	6 project manager's reports to Steering Group	no public access
Report	WP12	Benefits report	<a href="http://bit.ly/19V1rWS">http://bit.ly/19V1rWS</a>
Report	WP12	Final Report	<a href="http://bit.ly/15nbZHS">http://bit.ly/15nbZHS</a>

## 3.2 Activity and Learning

### 3.2.1 WP1: Audit current UH good practice data management

We used a variety of structured and unstructured methods, online and face to face, to investigate RDM practice in three main areas: health and human sciences, science and technology, and the humanities. Originally planned as an electronic audit in the first months of the project, we quickly discovered that one-to-one engagement was needed to elicit the complex factors at play with regard to RDM. The audit morphed into a continuing programme of engagement with individuals and workgroups. By the second year of the project researchers were coming to the RDM team with questions, but in answering, we continued to learn about practice and requirements throughout the entire project.

Audit activity consisted of:

- DCC mini-Cardio quiz sent to (30) likely senior researchers and leaders
- Full DCC CARDIO (Collaborative Assessment of Research Data Infrastructure and Objectives) attempt, sent to (20) senior researchers c-opted as a formally constituted stakeholder group
- DAF (Digital Asset Framework) survey broadcast via email to approximately 600 active researchers. The results from 67 responses are available via the outputs table.

- Extended conversations, directed by an interview protocol with 60 researchers.
- Workshops, presentations and consultations with individuals and at research groups and research institute meetings (approximately 300 people).

### **WP1 Learning:**

By an extrapolation from the survey data and with a 100% margin of error we estimate that there are **2PB of research data** in the University. This is about 20 times our existing available central storage capacity.

80-90% of the data is in the hands of well resourced STEM research groups. The remaining 200 - 400 TB is held by non-technical, less well resourced researchers (the so called long tail of research).

A lot of the data is working data or copies of data that need not be retained. Never-the-less, there is a significant body of data that should be kept and made available for reuse.

If even only 10% is retained we need to find ways of keeping 200TB of essentially dormant data, which must remain be accessible.

There is some good practice and generally good awareness with regard to safeguarding data. However, even where there is awareness and good intentions, practice can be less than robust.

Aside from a few large scale facilities, most data is held on workstations and laptops, and local ad-hoc storage. We find researchers trust devices that are liable to failure or loss, and that data disappears when the storage devices are 'archived' onto the shelf or under the desk, or fall into disuse because of system incompatibility.

There is significant use of insecure media, mostly USB sticks. There is significant use of unregulated, 'free' cloud services, particularly DropBox. The reason for using these media is that they facilitate sharing.

This results are our survey are consistent with the findings of other JISCMRD surveys and this suggests both the picture and the scale of the problem are reliable.

This work shows there is need for RDM support throughout whole project lifecycle, from data management planning, to safekeeping and collaborative working with data, to curation and arrangements for data re-use. The key points in a gap analysis reveal:

- a lack of awareness about, trust in, or inclination to use university services;
- a lack of recognition of data as a public, institutional or career asset;
- need for training and advice about all aspects of RDM;
- need for better facilities for collaborative sharing of working data;
- need for knowledge and facilities in regard of long term data preservation and re-use.

### **3.2.2 WP2: Cloud storage pilot (Health sciences)**

Over the course of 18 months we investigated a wide variety of products, services and potential solutions. The direction of these investigations was driven by the market; by what we knew about our researchers' practice; and by the overarching strategy to find cloud storage that would compliment existing facilities by extending functionality and/or capacity on more agile terms than were possible on campus.

We looked at product factors such as cost, usability, maturity, and terms of use; internal factors such as fit with UH data policy, fit with technology strategy, and ease of purchasing;

and usage factors such as sharing, group working and bandwidth considerations in wide area networks.

We invested time in understanding how our existing storage infrastructure might be better employed as part of a hybrid-cloud solution (a mixture of on campus and off campus networked facilities). We worked with the Thor Datacentre in Iceland (via our association with Herts Regional College) who provided facilities for a lot of our early trials and tests.

We installed DataStage, to assess its potential as a cloud deployable workgroup file share. In the same context we looked at iFolder, SparkleShare, OwnCloud, Microsoft SharePoint and FTP services.

We built an instance of ZendTo, which is an opensource system for transferring large files over the web. The requirement came from researchers but allowed us to test the integration of a cloud hosted system with UH systems.

In the context of security, and in response to demand from researchers, we assessed personal encryption solutions for transferring data in the cloud.

We used previous experience with running small group facilities and an incident of data loss to highlight the high cost of maintaining ad-hoc solutions favoured by individuals and small groups of researchers.

Although the core engagement in this work package was with researchers from health and human sciences, we also worked in history, engineering, physics and astronomy and there was a lot of overlap with other streams of work within the project.

## WP2 Learning:

We began this work expecting to find a portfolio of commercial cloud services that would slot into and compliment our existing infrastructure. This was naïve, perhaps, and proved not to be outcome. We have had some limited success in deploying cloud services but the main benefit has been in the learning, which is considerable, and in the main relevant beyond University of Hertfordshire.

By working with Researchers and from the evidence accruing in WP1 it was clear **that the predominant issue in working data management at UH was that of work group storage**. Robust storage, with good access, ease of sharing facilities, adequate capacity, and autonomous management, was the goal.

We found that our existing networked storage (our local cloud) could provide this facility to the majority of researchers, excepting only those with multi-terabyte requirements. The problem was that because of poor documentation, induction procedures and onward training most researchers couldn't use it effectively, thought it too limited in capacity, or were simply unaware of the offer. In most cases, these were perceived flaws, rather than real obstacles, and we can work to address this.

Given what we knew about the use of local ad-hoc storage devices we looked to the commercial cloud as the source of an alternative solutions. The problem here was work group functionality. This falls between the disparate offers of cloud files and cloud infrastructure. For example, RackSpace CloudFiles is very attractive and usable product but it is limited to one user per account. Whereas RackSpace CloudServers with elastic cloud storage can be configured any which way, but requires an experienced system administrator to do so. (RackSpace have a multi-user storage offer, but it is not yet as mature as the two services mentioned above, and is only available via their datacentres in the US).

We examined why Dropbox is so pervasive: it is simply a level magnitude easier to use than anything we or anybody else can offer. (Microsoft SkyDrive and GoogleDrive offer equivalent functionality but



this is encumbered and distracted by their inclusion in larger suites of services.) We also found this usability comes at considerable cost in terms of data replicated on every client machine (and backups thereof) and an unregulated channel from behind our firewalls into a system designed to share with any and everybody!

The search for an alternative, cloud hosted, 'DropBox like' solution was not productive. DataStage, iFolder, SparkleShare, OwnCloud were all too immature and unfinished to seriously consider (though the latter has been developed further since we looked at it, and forms the basis of an offering from the Australian National Data Service).

The latest 2013 Microsoft SharePoint/SkyDrive as a service looks like credible candidate for the work group, especially as it promises to integrate with local authentication, but may still be too complex an offer for many groups.

We also looked at the network and the movement of data around it and made several conclusions about the use of cloud storage.

Inside our network, network latency (the delay in transfer) makes our own storage, which is 'nearby' much faster than cloud storage. However, at any location other than on our campuses, the lower speed of the network equalises this situation and the performance for the user of our own servers and cloud storage is comparable. The way remote files are used matters too: transfer protocols such as SMB/CIFS and HTTPS/WEBDAV are talkative and involve a lot of handshaking which applications don't deal with well. So it is generally quicker to move a file to the point of use and open it, than to open it directly on a shared volume.

In a network related, but tangential issue, we found another barrier to using the cloud. Almost without exception cloud services charge for use of your data. For example, upload to RackSpace CloudFiles is free, but download costs £0.08 per GB. This is a difficult business model to work with in project work, which is funded by a fixed cost model.

The cost of storage is not easy to pin down. We estimate own datacentres cost £500-600TB/yr to run. Cloud storage costs £800 -£1100TB/yr plus egress charges. Archive storage is less expensive at £300-400TB/yr. The use of ad-hoc storage on (or under) the desk is attractive because it costs only £100-200TB/yr, but suffers very low fault tolerance and a low mean time between failure; with costs rising to £1600TB/yr in the event of a fault, or >£4000TB/yr in the event of data loss.

Although cloud storage looks expensive at first sight, on aggregate it may be less expensive. The problem is that research storage is not aggregated and remains dispersed and the costs of failure are mostly hidden and not shared between researchers; so the benefits of 'expensive' cloud storage are lost beneath the culture of 'it wont happen to me'.

The progression through all the issues above led us to a good understanding of the hybrid-cloud model and the disposition of our resources in it. Our Infrastructure proved to be big a ship to turn, even in 21 months. We haven't 'operationalised' very many new services as yet, but there is clarity about what we must do and how to move forward. We believe the following thesis will apply in many HE institutions:

- a hybrid cloud, made up of local university owned resources and remote leased resources can be effective in both functional and cost terms;
- the cloud is useful where lower performance is acceptable due to the application, or the network, or the frequency of use;
- a hybrid cloud should include storage with tiers of performance with active file management to move data between tiers according to demand
  - essentially: data should be placed where is it most frequently used; put data which is used on the web in the cloud; keep data which needs faster or frequent access nearby
  - there are many applications and scenarios where leased cloud storage is both appropriate and cost effective

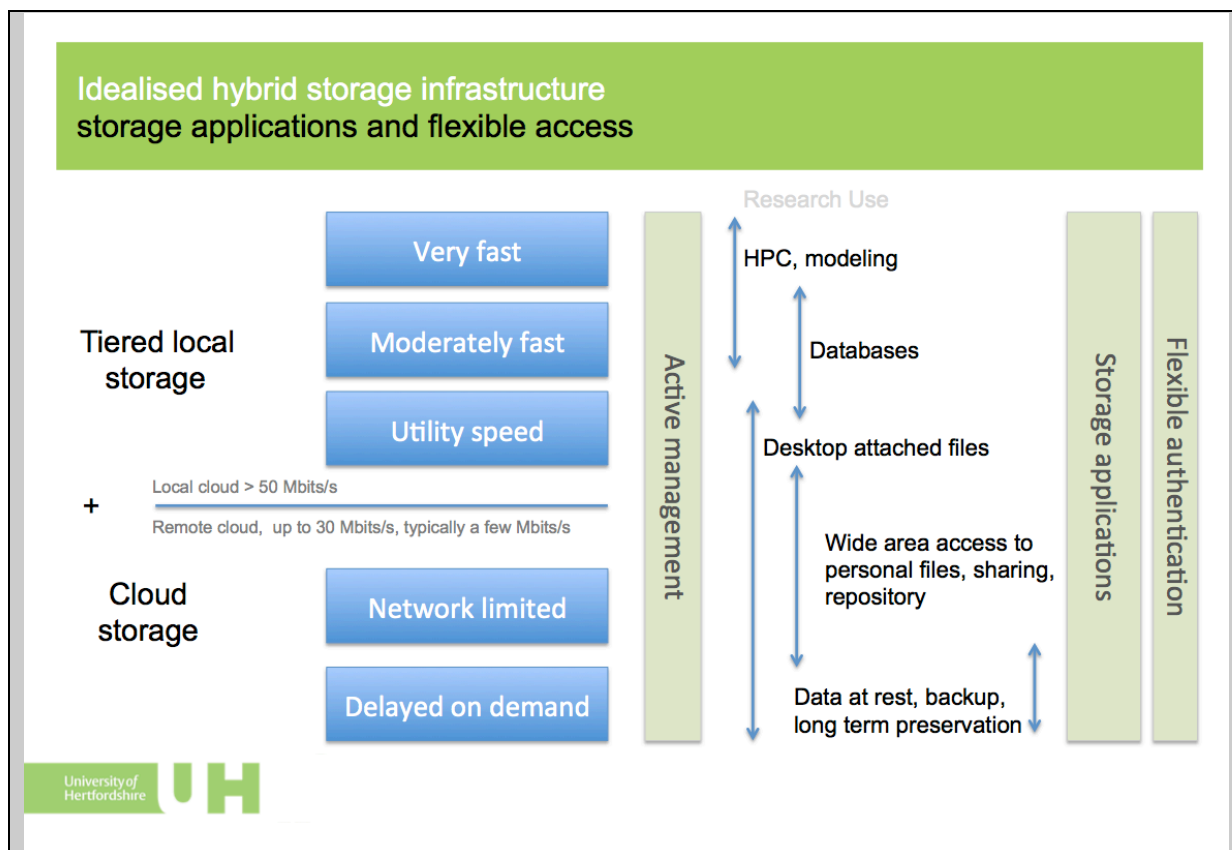


Figure 1 idealised hybrid cloud storage infrastructure

We found some very practical examples of the appropriate use of cloud storage which met known needs at UH.

The first was backup. Backed up data transfers asynchronously (unnoticed by the user) and is hardly ever used, which makes it suitable for remote storage. We tested a backup application during our work with Thor and found it worked seamlessly and effectively.

Another example was Zendto: a system for exchanging large files over the web. This integrates well with our local authentication and email systems, sits well on a cloud server where capacity can be expanded or contracted as required, and satisfies many of the demands for data sharing (or rather sending and receiving) from our researchers.

Another example is UHRA, our institutional repository, whose audience is in the cloud and which will contain large amounts of data that will be at rest. This is a perfect candidate to explore the use of very long term archive storage in the cloud of the kind offered by Arkivum A-stor.

### 3.2.3 WP3: Document Management pilot (Health sciences)

In this work package we set out to find out if the University's document management system (DMS) could be useful for research purposes. We use OpenText LiveLink, which is an enterprise Electronic Document and Records Management system, and is used to manage many of the documents associated with the business of running the university. It has all the familiar features of a document sharing platform such as Microsoft Sharepoint, with additional features for long term preservation, location management and disposal of both electronic and physical media.

We worked with the [Centre for Lifespan and Chronic Illness Research \(CLiCIR\)](#) who conduct research with sensitive, highly regulated data including clinical trials. In a clinical trial every document that is used to conduct and report on the research is stored in a Trial Master File, which is usually manifest as a set of locked filing cabinets in a secure room. The thinking was that the DMS could provide a secure and robust alternative to this.

We created an electronic Trial Master File (eTMF) by scanning and loading the TMF data from a trial that had recently completed. To organize the resultant 2000 documents we developed a File Plan, starting with the advice in the JISC Business Classification Scheme (BCS) and Records Retention Schedule (RRS) infokit, and adapting it to suit the material.

Having completed the first draft file plan we consulted with various other research groups across the University to come to a consensus of terminology, and a structure which was widely applicable. This gave us a generic DMS template for a research project which we rolled out to three other research groups for further assessment.

In addition to the file plan we considered what extra metadata was required to identify and discover a research project in the DMS. We also worked on a retention policy for research project data.

Finally we developed a workflow and form to support the increasing number of requests for 'storage' that this work was generating.

### **WP3 Learning:**

The University's document management system (DMS), hitherto unused by researchers, is appropriate for use in many circumstances.

The DMS is not an appropriate tool for storing large amounts of already structured data, but it is a good tool for recording the conduct of a project and or when a project uses common desktop formats to store data, particularly when versioning is required

The DMS becomes the tool of choice when the nature of the work requires a very high standard of file management and retention, subject to audit. This is often the case, for example, in health research.

The exercise of scanning and organising 2000 documents from a legacy health project helped us develop a generic file plan for a research project and also convince a number Principal Investigators of the efficacy of the system in their discipline. These PIs became advocates and demand for the system from throughout the Health and Human Science Research Institute (HHSRI) has been very high.

The file plan, instantiated as a DMS template, can be rolled out easily and although its structure is complex we can arrange the login point to be at the most appropriate folder for a project's needs.

With 10 projects online, 10 more engaged and many more probable this shows that the researchers' reluctance to engage with central services disappears when they are offered a tool which is superior to any other they have at their disposal.

In the wider context beyond UH this demonstrates the utility of MS SharePoint, though without the long term storage management features of LiveLink.

### **3.2.4 WP4: Thirty party repository pilot (Physics)**

The original intention of this work package was to work with the Centre for Atmospheric and Instrumentation Research (CAIR) to identify a dataset for deposit in an external repository,

and explore any issues peculiar to this data journey. It transpired no suitable dataset was forthcoming.

#### **WP4 Learning:**

A dataset was not forthcoming because our data was often already committed to the NERC British Atmospheric Data Centre, where the submission workflow is well established; or data was subject to collaboration agreements with several third parties, making it difficult to license for open access.

We did learn that astronomers deal with significant volumes of data (1-3TB each) during a project. Researchers who focus on simulations and modelling can produce up to 50TB of data during a three-year PhD project without keeping simulations from the development stages. It is not currently possible to retain this data. Similarly, our atmospheric physicists from CAIR create, use, and have to dispose of huge volumes of data during their work. Both groups have their own networks of local storage and share 400TB on the STRI HPC cluster. These are the best facilities by far in the University, but are still inadequate.

The raw data used by our Physicists and Astronomers is usually kept by the instrument operator and so does not need to be retained at UH. Some derived data is published, almost always by independent means by the PI. There is subject repository, the Strasbourg astronomical Data Centre (CDS) but this accepts only tabulated data which support a publication.

### **3.2.5 WP5: Applicability for the Humanities (History)**

History represents a constituency of researchers who are unflattering categorised as the 'long tail of research'. These researchers operate in small groups, have little technical support, and have data which is distributed through a variety of weakly structured formats. Their endeavours are individually small in scale but collectively significant at University of Hertfordshire. We undertook a programme of engagement with various History researchers and attended research group meetings and individual consultations in order to investigate their practice and see how RDM could be applicable in their discipline. The RDM Champion for the Social Sciences Arts and Humanities Research Institute (SSAHRI) also extended this engagement into other disciplines.

#### **WP5 Learning:**

We found many researchers who knew they should backup and safeguard their data but little or no knowledge of how to do this robustly.

Some individuals within the History group at UH are in the vanguard of the digital humanities (cf Old Bailey Online) and this culture was evident with all the Historians we engaged with: they were often keen to publish their data, but did not know how. One researcher had used the Historical Data Service before its demise. We were offered 4 oral history datasets to use test data for the development of UHRA.

In contrast to the Historians' enthusiasm for open data we found other researchers were possessive and hoarded their research: 'I haven't finished using it yet'; 'someone might use it and publish on it before me'. These 'reluctant publishers' highlight the need for RDM policy makers to find a better means of demonstrating that datasets can enhance a researcher's research profile as well as contributing to the public good.

All of the above evidence strongly suggests the projects outputs are needed in the Humanities and Social Sciences, where both working data management and the cultural of sharing need to be facilitated.

### **3.2.6 WP6: Review data protection and IPR issues**

This was a short work package which investigated how the University's position with regard to its ownership of Intellectual Property could be aligned with its commitment to publish open data,

We undertook a review of the licences commonly used for 'open' access to research information. The review took in Creative Commons, Open Data Commons, Open Government License, UK Data Archive licence and others. The main intention was to find a license for University of Hertfordshire datasets that was recognisable to the research community but consistent with our data management policies. We considered and took part in discussions about how the choice of licence might conflict with the prevailing definition of open data.

We also looked at the terms and conditions of RackSpace, who are a cloud storage vendor with a UK subsidiary.

#### **WP6 Learning:**

The license which meets the terms of our data policy and IP position is Creative Commons - By Attribution - Share Alike - Non-commercial (CC-BY-SA-NC)

We are aware that the community think this license is both impractical and against the spirit of open data, but we were unable to resolve the matter. The issue will be taken up by the University's Open Access working group.

RackSpace's UK terms and conditions proved acceptable to the Chief Information Officer, with the result that we able to use their cloud services and provide them to our researchers. The key factors were: provided you don't use their global content delivery network (which automatically caches data through the world) data is hosted in their London datacentre and thus subject to EU jurisdiction and law; and there were unequivocal and favourable statements about access and confidentiality.

### **3.2.7 WP7: Review long term sharing, storage and access issues**

This work package was rolled into a reconfigured WP8.

### **3.2.8 WP8: Data Repository (internal/external partners)**

Originally entitled 'Secure sharing pilot (internal/external partners)' the direction of this work package was changed when the it became clear that an institutional data repository would be required to fill the gap between national or international repository provision and the near term need to retain more data.

Information Hertfordshire already maintained the University of Hertfordshire Research Archive (UHRA), which is a repository of our traditional published work, based on DSpace opensource software. This was revamped and redeployed on upgraded hardware in order to make it fit for purpose to host data.

We also explored workflows for data deposit including the use of using SWORDII as the deposit protocol.

We also considered access protocols for data which is not entirely open and subject to permission criteria, and built a draft workflow for deposit and access to them.

We contributed to discussions about the appropriateness and practicality of using DataCite Digital Object Identifier (DOI) persistent IDs as an aid to citation and discoverability.

Our Current Research Information System (CRIS), which is an instance of Atira's PURE software, and is deployed in 20+ UK HE institutions, is being assessed as a catalogue of data, within and without the University. Work on this continues.

Toward the end of the project we re-evaluated the strategy to 'build out' UHRA on our existing infrastructure. We looked for a more sustainable system which supports infrequent access over the very long term. We considered Amazon Glacier and Arkivum A-Stor cloud based archival storage. We selected Arkivum and @mire (DSpace developers of Dryad) to develop our institutional data repository. This work, known as datasets@uhra, will continue after JISCMRD.

### **WP8 Learning:**

We found the existing version of UHRA was under powered, limited in capacity, and running an old version of DSpace on our first generation virtual machine (VM) infrastructure. It was necessary to upgrade all these aspects of the system. The system now runs DSpace 1.8, on a new, well resource Linux VM attached to elastic storage.

We tried, but failed, to deposit data via an instance of DataStage. DataStage did not develop as we had hoped during the lifetime of the project so this work was curtailed. SWORDII deposit was successfully demonstrated by @mire as part the UHRA upgrade.

Our discussions with British Library/DateCite indicate that we can demonstrate the necessary credentials to acquire DOIs for data (when our repository work progresses further).

Although we are committed to the concept of open data in the repository we found circumstances when the data is 'not quite open' but in which it can be published subject to some criteria. We drafted a workflow for the deposit and subsequent use of these exceptional data. The workflows, known as data journeys, cover open data, embargoed data, and restricted with access criteria data. This latter journey represents a problem, since the data is likely to be retained after its originators tenure at UH, so access criteria well need to be very well specified.

We also considered the extra metadata required for describing datasets. It is likely we will adopt the schema developed by our JISCMRD colleagues at Essex.

As the work progressed we began to discover potential datasets. We were offered 4 oral history datasets of WAV audio files, estimated at 200GB. This was 10x the size of the existing publications repository, and a reminder of the issues of capacity that we faced. We also noted that UHRA currently uses our costly high performance storage for data that is infrequently used and consumed mostly via the Internet. A better solution was needed.

When we looked for low cost, very long term storage, we selected Arkivum A-Stor. DSpace has been built over Amazon Glacier already, but Arkivum offered us very favourable terms, and were keen to collaborate with our DSpace developers (@mire) to build a functional proof of concept. A-stor also has the advantage of a direct connection to the JANET network and does not charge for data retrieval.

Discussion with JISCRMD colleagues at Leeds and Southampton reveal they are working with Arkivum to use A-Stor with ePrints, and we also have close synergies in the paths of our data journeys. We intend to take this work forward in collaboration as a result.

### **3.2.9 WP9 and WP10: The Research Data Toolkit**

This microsite contains 100+ pages and PDFs of RDM guidance. It covers all stages of the research data journey throughout a research project. This incarnation is the start of an

evolving but permanent resource which will be maintained by RDM professionals within Information Hertfordshire at the University of Hertfordshire. It is available at

<http://www.herts.ac.uk/rdm>

The site was produced as a joint effort with a sibling JISCMRD project, Research Data Management Training in Physics and Astronomy (RDMPA). It was written by researchers for researchers, and much of the material has been tested on UH training programmes, which it will support.

### **WP9 and WP10 Learning:**

For reasons expressed in much of the learning above the ToolKit was much less straight forward to construct than we had hoped at the outset. However, simply having the toolkit - which implies a practical and comprehensive resource - as an objective was a useful overriding steer for the whole project: we investigated, planned, and intervened so that we could eventually write good advice.

This advice alone is not an effective method of spreading the word. Training sessions, advertised throughout the University will be critical to the success of the ToolKit as an Agent of Change. We have developed training sessions, based on ToolKit materials, focusing on how to use RDM tools for project planning (DMPonline), security (TrueCrypt), and sharing (DMS and the UH shared drives).

We recognise that looking after working data and then preserving and publishing it are separate disciplines but the former is much easier to understand and 'sell' and than the latter. So we will leverage the former, exposing the ToolKit via a poster campaign using shocking imagery of damaged data and storage at risk, so as to create a foothold for engagement in the latter.

Although *Research Data ToolKit* served as a better working name for the project than *Service Oriented Toolkit for Research Data Management*, the word toolkit was perceived as too technical (rather than practical) in parts of our audience. For this reason, and to align with the practice of other JISCMRD projects, we will brand the Toolkit using *Research Data Management*, which works more generally to convey activity, service or discipline.

### **3.2.10 WP11A: Programme Engagement (external)**

Twelve staff were directly committed to the project in some part, at some point in its execution. All but one of these people took part in at least one JISCMRD or closely related event.

We attended a total of 67 of days of project related events in all. Of these 41 days were JISCMRD programme events, and almost all the rest were DCC events. The project manager accounted for 28 days. We gave 14 presentations or posters, mostly at JISCMRD workshops, but including the Association of Research Management Administrators (ARMA) Conference 2013 and the Royal Astronomical Society National Astronomy Meeting 2013. We estimate the audience for our presentations to be 480 with perhaps ~300 individuals.

We published 39 blog posts, 50+ tweets and made several contributions to [jiscmr@jiscmail](mailto:jiscmr@jiscmail) and [research-dataman@jiscmail](mailto:research-dataman@jiscmail).

### **WP11A Learning:**

DCC roadshows, RDMF events, and the BL/Datacite series were effective methods of acquiring knowledge, but the JISCMRD programme workshops and special workshops gave the most insight, and the project was both enlightened and directed by them.

The project also benefitted from the opportunity to make presentations at JISCMRD events, which often distilled our thinking, brought clarity to the issues, and allowed us to disseminate our work.

After a sceptical start the project manager came to know and enjoy the blog as a method reporting. We, sorry I, found the informal voice, the allowance of opinion, and human traits in writing can add to the authority of what is said rather than diminish it. However, I understand why some of team found it difficult to blog - the first person requires confidence (or perhaps careless abandon) to adopt.

### **3.2.11 WP11B Programme Engagement (internal)**

We took every and any opportunity to engage with researchers and service providers at UH. We estimate this involved nearly 500 interactions with between 200-300 people. In the main we used existing pathways and meeting agendas to infiltrate the research culture rather than try to make our audience come to us. We were good at bothering people, but always with the offer of help or a constructive message.

#### **WP11B Learning:**

What we learnt about RDM practice and the disposition of our resources to support it is covered throughout this report. The biggest lesson we had from the practitioners of research was: 'make it relevant to me', whether this relevance be expressed in terms of discipline specific issues, improvements to day to day working life, or career opportunity. Policy and an abstract public good do not appeal to already busy researchers if you are asking them to do something new or different. To slip into programme jargon to enforce the message: *you can make the pile of sticks as high as you like but there has to be a carrot underneath it.*

The programme wide attempt to gather evidence of tangible benefit acknowledges this, but it is worth re-iterating the need to demonstrate that public, institutional, and individual career benefit can be aligned.

### **3.3 Immediate Impact**

The project has brought an awareness of RDM as an issue to a large section of UH's research community, very few of whom were aware of it before our work. We know they were unaware from the answers given in our survey in 2012. We know this awareness now exists because we have spoken directly to hundreds of people via our programme of internal engagement. This engagement ranges from new post-Graduate researchers to the Chair Elect of the Board of Governors. We have been received with interest and attention by Principal Investigators (who will carry much of the burden of RDM) and research leaders alike.

RDTK has had the biggest impact on, and via, those researchers who were inclined to be helped and came to the project with a pressing issue. It has been effective mostly in the area of data management planning and working data management.

Our promotion of DCC DMPonline has produced 10 data management plans, which we will use in further advocacy. The Research Grant Office has adopted DMPonline and is using it to assist Principal Investigators with grant applications.

Our technical interventions have begun to deliver benefit in the form of reduced risk. Examples of this are: 200TB/70core HPC cluster moved from an insecure environment into our safe, green, cost effective data centre; 60 people have been trained in the use of



TrueCrypt to secure data on portable media; 2TB of health data have been moved from high risk desktop storage, to virtually nil risk cloud storage.

Our outputs are having impact that is evidenced by demand.

The work with health and human sciences has resulted in a high level of demand from that area for the improved University networked storage and document management offer. 20 research projects are engaged, whilst the entire portfolio of the Research Centre in Primary and Community Care (CRIPACC) plan to take up the offer.

The test phase of UHfileExchange, our new file sharing facility, has received good feedback and new users have made approaches by recommendation and word of mouth.

There is a waiting list for our next encryption workshop and a request for an expanded role in Generic Training for Researchers.

Insight into our existing facilities, and their context in the hybrid-cloud model, which was acquired with the help of our technology leaders, has impacted by making the case for tiered, active storage in the 2013/2014 technology plan.

The project has been effective at making new relationships: between technology providers and librarians within Information Hertfordshire (IH); between IH and the Research Grants Office; and between IH and research active staff. This is evidenced by the composition of the RDM team, joint presentations by each of these pairs, and invitations extended to the project to speak at many researcher forums.

It is hard to judge our impact beyond UH, but we believe we have made a difference to JISCMRD. Our blogs, which have focused on practical issues of interest to the whole programme, have been popular, with 2500 page views. In the latter part of the programme, RDTK presentations at workshops have shown that a research aspirational (rather than intensive), post 1992 university can make a significant contribution to the collective effort. Our discussions with organisations such as British Library, and in particular with the JANET brokerage (about storage in the cloud) have influenced their thinking.

### **3.4 Future Impact**

RDTK derived services and applications, and demands thereon, will see an increased utilisation and return on investment on several systems, including Storage Management Systems (R: drive), the Document Management System (DMS), the repository (UHRA), and the current research information system (RIS).

UHfileexchange promises to satisfy the demand to share research data in a great many cases and divert practice away from unregulated sharing arrangements such as Dropbox; as will more capacious, usable and sharable networked storage.

Active file management and Cloud storage applications such as Backup and Archive Storage for the repository will move data to the most appropriate and cost effective medium, and free up high performance storage for high performance use.

The impact of all these systems will be easily tracked via holdings and usage statistics. If usage increases benefit will be implicit: RDM will be have been made easier, more cost effective and less risky. There is a possibility demand will outstrip supply of central resources, but that will be a demonstration of success and in turn promote renewed innovative thinking in storage provision and retention policies.

The development of datasets@UHRA will impact in several ways: it will deliver a tier 3 (institutional) data repository; it will give a vehicle and focus to the issue of data re-use which needs to come to the fore in the near term; it will demonstrate a cheaper and more widely applicable alternative to existing arrangements for storing data at rest; it will contribute, via collaborative effort with other universities, to the body of knowledge about repository/storage integration and data journeys.

The RTDK legacy and project outputs will be promoted to successive cohorts of post graduate and incoming early career researchers via GTR and CPD. As these cohorts progress, RDM will mature and embed further, and the barrier of cultural change will diminish in favour of a recognisably necessary and matter of course professional discipline.

## 4 Conclusions

There are many conclusions that could be drawn from the project. These are the headlines:

- JISCMRD has been a success at UH.
- The RDTK project has made an impact in awareness raising and service development, and made good inroads into professional development and training. There are good materials, a legacy of knowledge and a retained group of people to sustain and develop the learning.
- We believe the service orientated approach shows that better technology can facilitate better RDM and the project has been an effective Agent for Change.
- We also understand that advocacy and training are as important as technology to bring about cultural change.
- Funding body policy and the implications of the ever increasing volume of data are understood. The business case is clear: the University cannot afford not to invest in RDM.
- JISCMRD phase2 has been an effective vehicle for knowledge transfer and collaboration. It provided an environment in which a new and complex discipline, and the many, interacting, conflicting, seemingly endless issues therein, could be explored with common cause and mutual support.

## 5 Recommendations

JISCMRD activity should continue, and try to reach the part of the research community that is least able to adopt RDM best practice without assistance, and won't do so as a matter of course. A profitable strand for JISCMRD3 would be Collaborative Services. Appropriate services would include joint RDM support services, or shared specific services, such as regional repositories (including DOI provision) or shared workgroup storage facilities. Institutions with advanced RDM capability could play a mentoring role. Another key strand would be Benefit of Data Re-use; to gather examples of innovative data use and academic merit and reward for individual data publishers.

The DCC should continue in its institutional support role. It should consolidate its DMPonline tool toward a cloud service, with features to allow organisational branding, and template merging. It should place new emphasis on the selection and publishing of data, with a

signposting tool for Tier 1 and Tier 2 repositories for subject specific data, including selection criteria, metadata requirements, and citation rates.

Opportunities for organisations to learn from each other and establish collaborations, which have been effective at JISCMRD2 workshops, should continue to be facilitated in some way. In addition, more attempts should be made to reach researchers directly in order to demonstrate the potential personal benefit of good RDM.

The JISC should continue to pursue national agreements via the JANET brokerage. These negotiations should be widened beyond Infrastructure as a Service to include RDM Applications as a Service (RAaaS), for example, Backup as a Service, Workgroup Storage, and Repository as a Service. The goal should be to achieve terms of use which satisfy institutional purchasing, IP and governance requirements; whilst allowing for acquisition by smaller intra-institutional units, from faculty, down to workgroup level. (JISC GRAIL- Generic Rdm Applications Independently Licenced) might be suitable brand for this activity. In addition, JANET should press cloud vendors for an alternative to 'pay-by-access' for data which is a barrier to uptake in fixed cost project work.

## 6 Implications for the future

JISCMRD2 has been conducted contemporaneously with the formation of a unified position from RCUK and other major funding bodies with regard to RDM. It is clear that robust safekeeping of working data, and the retention, wider sharing and re-use of derived data must be achieved if the University of Hertfordshire is to enhance its research position in line with its strategic plan.

The University of Hertfordshire is engaged in ensuring practical arrangements for addressing the demands and realising the benefits of RDM, the requirements of Open Access (OA), and the imperatives to demonstrate impact in research. From a strong base in pre-award support and strategy for the research community, we also now have a focus on developing institution-wide post-award support, which will likely increase subsequent to the 2014 Research Excellence Framework assessment.

Good practice RDM is recognised as a key element in continuing research excellence. The investment cannot be sustained at JISCMRD levels, but will be continued via both pre-award and post-award channels. The following roles will support RDM in the future: CIO, CTO, Head of Research Grants Office, RDM oversight and research systems group leader, Information Managers for Research, Research Grant Office advisors, RDM Project Officer, UHRA Repository Manager, Document Management and Systems Consultants. This amounts to a commitment of approximately 3FTE, not including senior management contributions.

The UH JISCMRD projects leave considerable momentum and several threads of ongoing activity, which will be sustained. These include;

- RDM training via Research Grants Office researcher CPD programme, Generic Training for Researchers (Post Graduate programme), Research Institute training
- Finessing and adding to the Research Data ToolKit at <http://www.herts.ac.uk/rdm/>
- A poster and event campaign to market the project outputs and the benefit of good practice RDM, leading up to participation in the induction of the 2013 cohort of post-graduate researchers.

- Data Management Planning support via Information Hertfordshire and Research Grants Office.
- Storage infrastructure improvements technology plan for 2013/2014.
- New 'Research Storage' offer (r: drive / document management)
- UH file exchange (working data sharing)
- Datasets@UHRA, (institutional data repository)
- Research Data Catalogue (metadata catalogue in Current Research Information System)
- Annual review and refinement of UPR IM12 Data Management Plan and RDM appendix
- Continued engagement in HE collaborations, and lobbying of JANET to secure nationally negotiated cloud services

## 7 References

References have been made by direct links to online resources where appropriate.

## Appendices

### **Appendix A - Programme Engagement (external)**

Attached as rdtk-final-appendixA.pdf

### **Appendix B - Programme Engagement (internal)**

Attached as rdtk-final-appendixB.pdf

## RDTK Appendix A - Programme Engagement (External)

person	event	location	JISCMRD days	Other days	Presentation	est audience	
Jo Goodger	RAS National Astronomy Meeting 2013	St Andrews	01/07/2013		5 "Preserving Digital Data at UH" to follow	50	
Bill Worthington	ARMA 2013 Conference	Nottingham	11/06/2013		1 rdtk-rgo-arma-conf-nottingham-2013.ppt	40	
Jo Goodger	JISC-BL DataCite Workshop (6)	London	14/06/2012	1			
Linda Wilks	JISC-BL DataCite Workshop (6)	London	14/06/2012	1			
Liz Nolan	ARMA 2013 Conference	Nottingham	11/06/2013		1		
Bill Worthington	Now and Future of Data Publication	Oxford	22/05/2013	1			
Bill Worthington	JISCMRD Funding workshop	Aston	26/04/2013	1			
Bill Worthington	JISCMRD Closing workshop	Aston	25/03/2013	2	rdtk-towardthehybridcloud-aston-2013.pdf	50	
Bill Worthington	JISCMRD Closing workshop	Aston	25/03/2013		rdtk-agentofchange-aston-2013.pdf	30	
Graham Brown	JISCMRD Closing workshop	Aston	25/03/2013	2	rdtk-dms-poster-aston-2013.pdf demo	15	
Mohamed Hansraj	JISCMRD Closing workshop	Aston	25/03/2013	2	rdtk-poster-aston-2013.pdf demo	10	
Jo Goodger	JISCMRD Closing workshop	Aston	25/03/2013	2	rdmtpa-poster-aston-2013.pdf	60	
Jo Goodger	JISCMRD Closing workshop	Aston	25/03/2013		rdmtpa-presentation-aston-2013.pdf	30	
Jo Goodger	DaMSSI-ABC Training Strand workshop	London	18/03/2013	1			
Mohamed Hansraj	JISC-BL DataCite Workshop (5)	London	08/03/2013	1			
Jo Goodger	JISC-BL DataCite Workshop (5)	London	08/03/2013	1			
Jo Goodger	P&A training discussions with Stephane Goldstein	London	04/03/2013		3		
Bill Worthington	JANET/JISCMRD Storage Requirements Workshop	London	25/02/2013	1	rdtk-storagereq-paddington-2013.ppt	60	
Mohamed Hansraj	CKAN for RDM in an Academic Setting	London	18/02/2013	1			
Linda Wilks	JISC-BL DataCite Workshop (4)	London	03/12/2012	1			
Bill Worthington	JISCMRD benefits evidence workshop	Bristol	29/11/2012	2	rdtk-jiscmr-bridol-benefits-november-2012-v4.pdf	20	
Sara Hajnassiri	JISCMRD benefits evidence workshop	Bristol	29/11/2012	2			
Mohamed Hansraj	DCC Roadshow London	London	20/11/2012		1		
Jo Goodger	DCC Roadshow London	London	20/11/2012		1		
Bill Worthington	Research Data Management Forum 9	Cambridge	14/11/2012		2		
Bill Worthington	JISCMRD Training Stand Launch Workshop	London	26/10/2012	1			
Jo Goodger	JISCMRD Training Stand Launch Workshop	London	26/10/2012	1	rdmtpa-jiscmr-notts-oct-2012.ppt	35	
Mohamed Hansraj	JISCMRD progress workshop	Nottingham	25/10/2012	2			
Bill Worthington	JISCMRD progress workshop	Nottingham	24/10/2012	2	rdtk-jiscmr-notts-oct-2012-storage-web.pdf	60	
Bill Worthington	JISCMRD progress workshop	Nottingham	24/10/2012		rdtk-jiscmr-notts-oct-2012-ds-workflow-web.pdf	20	
Bill Worthington	JISCMRD progress workshop	Nottingham	24/10/2012		rdtk-jiscmr-notts-oct-2012-ds-poster.pdf		
Mohamed Hansraj	JISC-BL DataCite Workshop (3)	British Library	10/09/2012	1			
Bill Worthington	Open Repositories 2012	Edinburgh	09/07/2012		2		
Mohamed Hansraj	JISC-BL DataCite Workshop (2)	London	06/07/2012	1			
Bill Worthington	JISC-BL DataCite Workshop (1)	London	25/05/2012	1			
Sara Hajnassiri	Meeting Disciplinary Challenges in Research Data Management Planning	London	23/03/2012		1		
Sara Hajnassiri	Institute of Historical Research, Preservation and research data: what's in it for me?	London	14/03/2012		1		
Bill Worthington	JISCMRD policy workshop	Leeds	13/03/2012	2			
Liz Nolan	JISCMRD policy workshop	Leeds	13/03/2012	2			
Bill Worthington	Dataflow Launch	Oxford	02/03/2012	1			
Sara Hajnassiri	DCC Roadshow London	London	21/02/2012		2		
Bill Worthington	JANET Brokerage Launch	London	09/02/2012		1		
Bill Worthington	JISCMRD Launch Workshop	Nottingham	01/12/2011	2			
David Ford	JISCMRD Launch Workshop	Nottingham	01/12/2011	2			
Bill Worthington	ViDaaS workshop	Oxford	22/11/2011	1			
Bill Worthington	DCC Roadshow Cambridge	Cambridge	09/11/2011		3		
Bill Worthington	Research Data Management Forum 7	Warwick	02/11/2011		2		
				41	26	14	480
					67		

## RDTK Appendix B - Programme Engagement (UH)

rdm staff	event	group	date	audience
Sara Hajnassiri	Digital Asset Framework Survey	600 researchers	Jun-12	67
Bill Worthington	History Group Consultation	UH	20/09/2012	5
Bill Worthington	HHSRI Research group meeting	UH	03/07/2012	20
Bill Worthington	IH Lunchtime Seminar	UH	07/02/2012	15
Bill Worthington	Karen Friedli consultation	UH	15/03/2012	1
Mohamed Hansraj	Uh Shared data meeting	UH	12/12/2012	5
Mohamed Hansraj	Truecrypt Workshop	UH	04/12/2012	22
Mohamed Hansraj	Truecrypt Workshop	UH	22/01/2013	18
Jo Goodger	Welcome new STRI PGRS	STRI	01/03/2013	8
Jo Goodger	Discussions with Tim Gledhill, training coordinator in Astronomy	STRI	03/10/2012	1
Jo Goodger	Discussions with Yasmin Imani, training coordinator for GTR	HBS	03/10/2012	1
Jo Goodger	RDM update	Staff Dev.	16/10/2012	16
Jo Goodger	Intro. RDM	CAR	23/10/2012	6
Jo Goodger	RDM discussions with Astronomy staff and PGRS	STRI	27/08/2012	20
Jo Goodger	RDM discussions with STRI staff	STRI	30/08/2012	6
Cathy Tong	RDM update	Staff Dev.	30/04/2013	10
Cathy Tong	Intro. RDM	GTR	13/05/2013	10
Graham Brown	Document Management Service Provision and Training	UH	Various	20
Graham Brown	Document Management External Users/Research Collaborators	Collaborating Universi	Various	11
Graham Brown	Document Management Demonstrations	UH	Various	30
Graham Brown	File plan development	UH	Various	19
Graham/Champions/Team	Additional Document Management Consultations	UH	Various	11
Sara Hajnassiri	Interview Richard Greenaway	UH	24/05/2012	1
Sara Hajnassiri	Meeting with David Welsted- RDM assessment	UH	29/06/2012	1
Bill + Sara	Presentation for Centre for Engineering and	UH	30/04/2012	6
Sara Hajnassiri	Centre 4 regional & local history research -	UH	12/06/2012	9
Sara Hajnassiri	PAM School meeting	UH	13/06/2012	15
Sara Hajnassiri	H&HSRI Research group gathering, RDM	UH	03/07/2012	10
Sara + Bill	Interview with historians	SSAHRI	20/09/2012	4
Bill+ Sara	Stakeholder forum meetings	UH	10/12/12 + 22/03/12	32
Sara Hajnassiri	DAF survey	UH	May-12	
Sara + Cathy	Lunch time seminar	UH	27/03/2012	12
Sara + Graham+ Bridget	RDM update for H&HSRI	HHSRI	17/05/2012	14
Bridget Russell	Health and Human Sciences researchers	HHSRI		20
Bill Worthington	Director CAIR + PI atmospheric physics		Nov-11	2
Bill Worthington	Director CAR + Reader Astrophysics		Nov-11	2
Bill Worthington	Director of Research STRI	STRI	Nov-11	1
Bill Worthington	Director of Research HHSRI	HHSRI	Nov-12	1
Linda Wilks	Director of Research SSAHRI	SSAHRI	19/02/2013	1
Linda Wilks	Associate Dean Research Humanities	SSAHRI	05/03/2013	1
Linda Wilks	Humanities: Early Language Intercultural Acquisition	SSAHRI	21/03/2013	1
Linda Wilks	Humanities: Remembering the First World War; Low Carbon Past: Heritage Hub	SSAHRI	20/03/2013	1
Linda Wilks	Humanities: social history	SSAHRI	20/03/2013	1
Linda Wilks	Associate Dean Research Law	SSAHRI	05/03/2013	1
Linda Wilks	Associate Dean Research Creative Arts	SSAHRI	27/03/2013	1
Linda Wilks	Creative Arts: film	SSAHRI	14/03/2013	1
Linda Wilks	Creative Arts: music	SSAHRI		1
Linda Wilks	Creative Arts: music	SSAHRI		1
Linda Wilks	Business School: Film, AV, Creative industries	SSAHRI	20/03/2013	1
Linda Wilks	Business School: Labour	SSAHRI		1
Linda Wilks	Business School: Event management	SSAHRI	21/03/2013	1
Linda Wilks	Business School: Operations management	SSAHRI	21/03/2013	1
Linda Wilks	Business School: Finance	SSAHRI	14/03/2013	1
Linda Wilks	Business School: Global political economy	SSAHRI		1
Linda Wilks	Business School: Marketing	SSAHRI	16/05/2013	1
Linda Wilks	Business School	SSAHRI	13/06/2013	1
Linda Wilks	Business School: Management and Strategy	SSAHRI	06/06/2013	1
Linda Wilks	Associate Dean Research Education	SSAHRI	14/03/2013	1
Linda Wilks	Education: Student learning	SSAHRI	15/05/2013	1