Project Identifier: RDMTPA
Version: 0.4
Contact: Dr William Worthington
Date: 5-Jun-14

JISC Final Report

Project Information			
Project Identifier	RDMTPA		
Project Title	Research Data Management Training for the whole project lifecycle in Physics & Astronomy research		
Project Hashtag	#rdmtpa_herts		
Start Date	1 June 2012	End Date	31 July 2013
Lead Institution	University of Hertfordshire		
Project Director	Dr David Ford		
Project Manager	Dr William Worthington		
Contact email	w.j.worthington@herts.ac.uk		
Partner Institutions			
Project Web URL	http://research-data-toolkit.herts.ac.uk/		
Programme Name	Digital infrastructure: Research management programme – RDM Training		
Programme Manager	Dr Simon Hodson		

Document Information			
Author(s)	Dr Joanna Goodger and Dr William Worthington		
Project Role(s)	Project Researcher and Project Manager		
Date	12 June 2013	Filename	RDMTPA-Final-Report-v4.doc
URL	http://uhra.herts.ac.uk/handle/2299/13638		
Access	This report is for general dissemination		

Document History		
Version	Date	Comments
0.1	29 April 2013	First draft for initial review by team
0.2	12 June 2013	Second draft for initial review by Stakeholders
0.3	20 May 2014	Final version for repository deposit. Updated links in Table of Outputs, including links to training materials on JORUM.ac.uk
0.4	5 June 2014	Updated with persistent identifier http://uhra.herts.ac.uk/handle/2299/13638

Table of Contents

1 Acknowledgements

The Research Data Management Training for the whole project lifecycle in Physics & Astronomy research – RDMTPA – project was funded by JISC as part of the Digital infrastructure: Research management programme, Research Data Management Training stand during the 2011-2013 programs.

Within the University of Hertfordshire (UH), we owe thanks to our stakeholders; Dr Jyoti Choudrie, Prof Bruce Christianson, Dr Ornella Corazza, Mr John Corkery, Dr Hulya Dagdeviren, Dr John Done, Prof Janet Drew , Dr David Ford, Dr Karin Friedli, Dr Richard Greenaway, Ms Sara Hajnassiri, Prof. Martin Hardcastle, Prof. Di Martin, Mr Andrew McHugh (from DCC, Mr Paul Nicholson, Liz Nolan, Dr Doug Prain, Dr Sooda Ramalingam, Prof. John Senior, Dr David Wellsted , Dr Bill Worthington, and Dr Andy Wroot.

2 Project Summary

Research institutes world-wide are facing dramatic changes to the way that data is preserved which is impacting their data management from the initial application for funding to the ways in which data are stored, shared and may public after publication. At the University of Hertfordshire (UH), we conduct world leading research across several disciplines, including History, Nursing and Midwifery, Engineering, Physics, Computer Science, English, and Art and Design. The previous JISCMRD programme focused on new resources particularly for the social sciences and humanities; however, the physical sciences were under-represented, perhaps because these disciplines are seen as more technical and by extension inherently likely to practice good RDM. This is only partially correct.

Within the physical sciences, UH operates the Centre for Astrophysics Research (CAR) and the Centre for Atmospheric & Instrumentation Research (CAIR). Both Centres conduct world-leading and internationally excellent research, and have responsibility for the sharing, dissemination and maintenance of large volumes of data. They also represent a large cohort of post-graduate students and early career researchers who will benefit from discipline specific training in research data management (RDM).

We have developed existing JISCMRD work and collaborated with researchers in CAR and CAIR to develop a short course in RDM for Post-Graduate and early career researchers in the physical sciences adopting a whole project lifecycle approach, from data management planning, through good data safekeeping, to curation options and arrangements for data reuse.

Beyond this initial remit, we have gathered generic best-practice RDM advice from other researchers across the university and developed an “Introduction to RDM” course as part of our existing postgraduate training programme: 'Generic Training for Researchers' (GTR) and our staff development program.

In addition, we have included packages of discipline specific examples, utilising the work of other RDM training projects and collecting additional information from the researchers at UH in the remaining research institutes (HHSRI and SSAHRI). These additional supplementary slides and guidance can also be used to train support and technical staff.

All of these materials will be published to the JORUM archive and submitted to the VITAE catalogue after evaluation of these resources this autumn.

3 Main Body of Report

3.1 Project Outputs and Outcomes

Output / Outcome Type	Brief Description and URLs (where applicable)
UH DMP*online* Template	Progression from a RDM checklist within our UH Data Policy, to a DMP*online* template that fulfils the UH data policy and stands alone as a record of the treatment and location of data.
Project Website	Including guidance on best-practice RDM for topics related to the lifecycle of research projects and the following training materials http://bit.ly/uh-rdm
Training Slides	Presentation slides covering 18 topics within four RDM sessions: 1 – Planning a project http://find.jorum.ac.uk/resources/18502 2 – Getting started http://find.jorum.ac.uk/resources/18503 3 – Safeguarding your data http://find.jorum.ac.uk/resources/18504 4 – Finishing touches http://find.jorum.ac.uk/resources/18505
Trainer Notes	Aims and key points for each slide of the training.
Discipline Packages	Examples to make the generic advice relevant in physical sciences; Physics and Astronomy. Also in Health sciences, History, and Business. (Additional packages to follow in the coming months.)
How to choose training	Advice on which training is suitable and how these materials can be used in training sessions for researchers, research students, support and technical staff within and without UH.
Case Studies	Descriptions of 12 projects, highlighting RDM practices, and key issues and solutions that have affected researchers throughout the university, posted on our RDM website for the benefit of other researchers in the university. http://bit.ly/uh-rdmcs
Current and best-practice assessment	Formal and informal interviews with researchers in Astronomy, Physics, Maths, Robotics, and Atmospherics to discuss the bespoke solutions they have adopted and the applicability of our RDM tools to the physical sciences.
Development Blogs	Blog summaries on • the progression from RDM training sessions for astronomers to generic training sessions for researchers in all disciplines, • the development of the website • the development of the UH DMP Template • http://research-data-toolkit.herts.ac.uk/
Evaluation of Training	Feedback evaluated after each training session used to improve training sessions in particular the content and duration.
Improved data management in astronomy research students	Follow up interviews with research students demonstrated improved awareness of data management, preservation requirements and security of data.

Workshop presentations	This work has been presented at JISC workshops and RDM training related meetings; 24/10/12 – JISC Building Institutional RDM Meeting in Nottingham *"RDM Training for Physics and Astronomy"* 26/10/12 – RDM Training Stand workshop *"RDMTPA at UH"* 25/03/13 – JISC RDM Meeting in Birmingham *"RDM Training at UH"*
Presentations for researchers	"Introduction to RDM" presented to researchers, staff and students. • Staff development: 16/10/12, and 30/04/13 • GTR: 13/05/13 • For Astronomy PGRS: 23/10/12 • For STRI new PGRS: 01/03/13 Research Group seminars are planned for the autumn term 2013. "Preserving Digital Data at UH" presented at the National Astronomy Meeting in St Andrews, 1-5/07/13.

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

We began our training project on the back of the RDM Toolkit project at UH (RDTK). The team had already issued a Data Asset Framework (DAF) questionnaire to the researchers at UH and received 68 responses - XX% of researchers at UH. Of these responses 47% were from researchers in the physical sciences, supporting our hypothesis that these disciplines are more technical and aware of their data needs and of general RDM. This also suggested that these researchers would be open to discussing their RDM and benefit from the tools being developed by our RDTK project. The DAF also revealed that training courses and written guidance were welcome as long as they are teamed with existing training and are not 'long-winded' (Fig. 1). It was also clear from these responses that face-to-face coaching and help with DMPs was important, particularly as DMPs as part of the funding process are new to researchers.

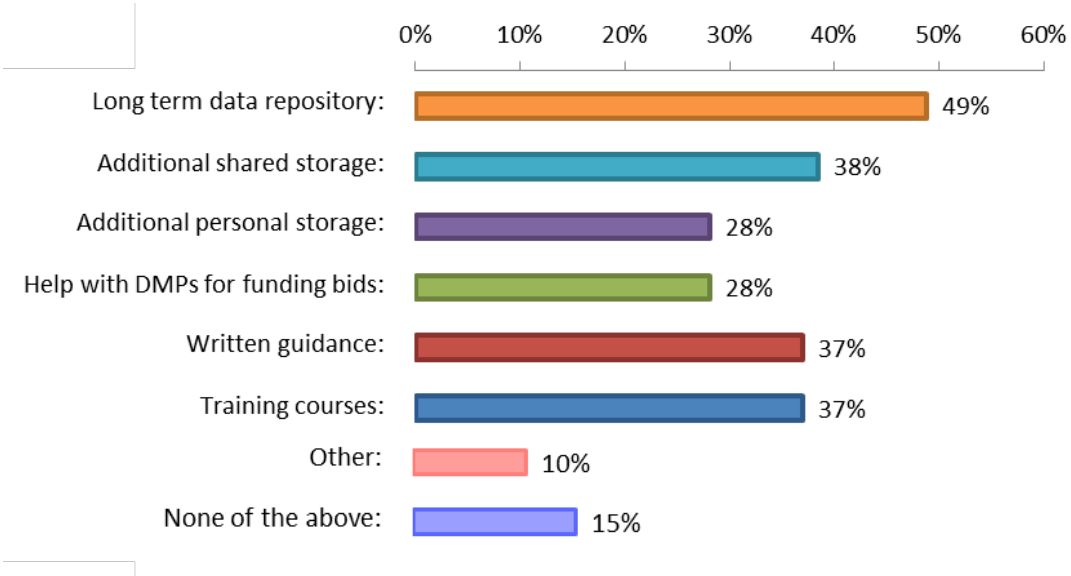

Fig 1: Results of the UH DAF survey, Q24 - What support would you find useful in helping you to manage your research data?

Encouraged by our researchers' responses, our training project progressed following the aims set out in our project plan, which were split into clear work packages and deliverables. The following sections describe the development of these work packages and how the evaluation and review of the deliverables resulted in the outputs and outcomes mentioned above.

3.2.1 Data Management Plans

The first challenge was to meet the new demands of funding bodies by echoing the requirement of a Data Management Plan (DMP) within our own institute. Our previous Data Policy included a checklist of issues that should be considered during the planning stages of a project. A data management plan was not compulsory and the checklist was repetitive and cumbersome. We have since clarified the UH Data Policy replacing the checklist with a requirement for researchers to complete the UH DMP available on the DMP*online* system.

We required the UH DMP template to be comprehensive, to be a stand-alone document, and not to require too many addition questions as this DMP would be completed simultaneously with the DMP template for the researcher's funding body. As our researchers apply for funding from all of the research councils in the UK, collectively known as RCUK, we compared the questions within these RCUK templates to our own checklist and determined which questions are most common and therefore most important to funding bodies. As shown in Table 1, our template does not include every question that is required by the funding bodies, but reflects the main features of RDM throughout the project lifetime;

- what data will be gathered and in which formats,
- what metadata will be recorded,
- where the data will be stored,
- where the data will be preserved,
- how access will be restricted for sensitive data.

Table 1: Comparison of research council DMP template questions to those in the UH DMP template, which includes 65 questions over 11 RDM themes.

	AHRC	BBSRC	ESRC	MRC	NERC	STFC
Total Questions asked in template	43	27	29	22	50	30
Questions in common with the UH template	38	23	23	20	38	22
Percentage of UH DMP completed by migration	58	37	37	31	58	34
Themes in common	7	7	7	8	9	5

Once the content of the UH DMP Template had been approved by our senior management, we conferred with our RDM champions within HHSRI and SSAHRI to develop advice to accompany the template on DMP*online*. They highlighted the fact that not all questions are directly relevant to all researchers. We cannot reasonably produce templates for each subject, nor can we accept the funding body DMPs for our records as they are designed to accompany comprehensive applications; they will not stand alone. This is clear from the missing themes in some of the templates; the STFC template does not include any questions about existing or new data as this is included in the body of the application. The template for MRC, BBSRC, and AHRC do not ask about ethics or intellectual property as this is also included elsewhere in their applications. We have also included 19 questions which put the report in context and discuss resources for reviewing the data and the DMP beyond the project life-time, which accounts for 29 per cent of our DMP.

3.2.2 Best practice guidance for working data

We began interviewing Astronomers within CAR to identify existing behaviour with respect to how researchers are gathering, storing, and sharing their data within UH and with their external collaborators. Our CAR research group includes 34 research staff and 34 post-graduate research students (PhD and MRes) and we

interviewed 20 researchers (29 per cent) including 7 post-graduate research students and 8 post-doctorate research staff.

These researchers work mainly on Linux and Mac operating systems, occasionally using Windows partitions or separate windows machines for designing presentations and posters in MS PowerPoint. As Windows is used infrequently, these researchers do not have access to the university shared drives or to the advance features of MS Outlook, which are used by the service staff to organise meetings etc. Instead, they store their data on their own machines or the STRI cluster, which is not backed up, and while they can use email clients for correspondence, the calendar is inaccessible.
The University of Hertfordshire uses *Novell* to network its staff machines and *Network Connect* to allocate staff machines UH IP addresses. This system can also be used off campus and on personal laptops to access existing UH services; however, as the IP is affected, this interferes with how Linux machines are set up. We have suggested the installation of Windows Virtual Machines (WVMs) to overcome this obstacle and allow access to our services and RDM tools. In particular, researchers can access the UH shared storage facilities; we offer shared drives for our academic schools and research groups, and the document management system (DMS).

Our interviews also showed that over half of the researchers questioned work on laptops that are not automatically backed up by the department, but they do back up to external hard drives at least weekly. Those with mac machines generally use *time machine* to back up crucial files hourly, daily, and weekly. Our conversations did highlight that storing their backup in the same bag, vehicle, or room still puts the data at risk – if someone is going to steal a laptop, they're not going to leave your back-up behind. Many researchers were not aware of the difference between using the networked and local drives on their UH desktop machines; this was particularly true of research students. We made them aware during these discussions that their networked drives are backed up automatically by the IT services in CAR, but their local drives should be used as an independent backup, that they should set up themselves using *rsync* and *cron* Linux tools.

As the researchers in CAR are allocated web space, files that are too large to share by email and are not sensitive are shared via the net, or using *DropBox* as it is openly accessible, easy to use, and available on all operating systems. With WVMs, the shared drives are available to these researchers as they are members of staff. Access does not automatically extend to research students or to external collaborators; however, we have now established workflows for research students (post-graduates, masters, and under-graduates) and external collaborators to gain visitor memberships to UH which will give them access to their groups' shared drives.

These formal and informal conversations with staff in CAR demonstrated a lot of good practice with regard to working data, and allowed us to start introducing the tools developed in the RDTK project to our researchers and in return, identify issues which would inhibit the use of the tools by all researchers. As the methods for storing and sharing data were similar throughout CAR irrespective of the research area they study, we extended our scope to include other researchers in the STRI. We interviewed members of staff from Atmospherics, Maths, Robotics (computer science), and Physics. Despite the diverse range of file formats and the group storage facilities, the RDM is very similar throughout the STRI and our tools are equally as useful in these research groups.

Surprisingly, we found that the security of sensitive data is an issue in the physical sciences. We had expected that this would be true in health and human sciences or in business, but atmospheric researchers develop instruments that may be pending patents, and the robotics research group develop robots that interact with children and vulnerable adults whose personal information and audio-visual interviews are sensitive. Our encryption tools have proved to be invaluable to these groups.

These issues; storage, sharing, back-up, and encryption, are the foundation of our working on data best-practice guidance. We have collated this advice and published it on our RDM website 'Working' pages and produced a 'Safeguarding Data' training package, which is available on the website and will be preserved in the JORUM archive and VITAE training catalogue (see §3.2.6 for more information on JORUM and VITAE).

3.2.3 Best practice guidance for preserving data

To conclude on the best practice guidance for preserving data, we first had to establish what the current requires were and how researchers preserved their data. In Astronomy, data is collected by ground- or space-based observatories and stored in their raw, unprocessed state at their facilities. Researchers request observations of target of interest and if successful, they have a proprietary period during which they alone have access to the data. After this period, the data is openly available to the community from the observatory archive.

Once processed and analysed, the measurements, simple images, and plots are published in papers and tabulated data may be published in the Strasbourg astronomical Data Centre (CDS). The processed data is not generally published. An exception is when the data is part of a large survey or project and the images are published online as a data release with a digital object identifier (DOI). These projects are usually international involving multiple institutes, and are likely to include one institute that has an institutional data centre or archive. In the near future, UH will offer institutional repository for our researchers, which will include a data catalogue of all published data related to research at UH. The current practice is to contact the lead author of the publication and request further information and/or data which are traded for collaboration, co-authorship or acknowledgements.

Many of the atmospheric researchers are funding by NERC, the Natural Environment Research Council, who require researchers to deposit their raw and processed data in the national NERC archive. However, any measurements or results from these data, including data used to produce plots, are not preserved in the NERC archive. They are either published in papers or articles, or stored locally. This highlights the fact that even if there is a national data archive for a subject, it may not include all stages of data, which demonstrates another aspect of RDM that would benefit from a UH archive. Some would argue that the processed data is the most important as it can be used immediately, while the raw data is crucial for comparison studies as the processing of data is not uniform between publications even when the same software and/or methods are used. The measurements and numerical results are usually published in the article, but shouldn't they be available electronically as well? This is certainly true for large datasets that cannot be reproduced manually.

Comparing the preservation practices of these research communities with others outside of the physical sciences at UH, we found that few researchers feel that they can share their data due to sensitivity issues, and many others do not have a national archive available. It was clear that producing our own institutional archive will benefit the majority of our researchers particularly as publication and preservation of data will be required by the RCUK. This has also meant defining metadata requirements and preservation guidelines.

As subject repositories have their own guidelines for which data should be preserved and the formats and naming conventions that should be used to archive these data, it is difficult to produce generic advice for all of our researchers. However, by applying some common sense and training our researchers to be prepared to preserve data, we have produced a 'finishing touches' training module that suggests researchers choose data that is likely to be reused by themselves and their community and also to consider which tools and methods could be useful to others. The preservation examples from our researchers, as well as discussions with members of the NERC archive and the Archaeology Data Service (ads), have enabled us to compile best-practice guidance on finishing projects on our RDM website that includes advice on publishing, preserving, archiving, and web publishing. These topics are also included in our 'Finishing Touches' training module, which will be published online and in JORUM and VITAE (see §3.2.6) once the UHRA accepts data.

3.2.4 UH Training courses

The purpose of this project was to produce training materials for Physics and Astronomy researchers focusing on post-graduate research students (PGRS). Combining the UH DMP template, which considers the whole life-cycle of projects, and the responses from the researchers that we interviewed in CAR, we delivered an hour training session entitled 'An Introduction to Research Data Management' to first year PGRS in Oct 2012 as part of the CAR training program. We provided a four-sided DMP developed specifically for PGRS which included simply-worded, basic questions, as the full UH DMP was still in development and was deemed too technical for

a PGRS project. We covered why a DMP is required and how it benefits researchers to publish their data, as well as an overview of the UH resources available to them; the shared drives, the DMS, the UHRA, and the CAR back-up policy. The feedback was good (averaged 4.25/5), the lowest goodness score relating to sharing their data. This was probably because the tools for sharing data were still under development and descriptions were intentionally vague. A full summary of the feedback from all of our training sessions is provided in the appendix with the DMP hand-out, which proved useful for relating the generic advice to their projects; however, a live website for future reference would have been of benefit.

A similar session in the staff development program aimed at researchers, who can expect to submit a DMP during a grant application, was well attended (16 attendees). Again, the feedback was acceptable (3.83/5) with the lowest scores relating to encryption. This session was held in a computer suite so that researchers could register and begin completion of a DMP*online* template. Many researchers felt that the course was too short and that more time should be spent on explaining the tools for storage, sharing, archiving and encryption.

In light of this and interviews with STRI researchers, we have since delivered two sessions in the staff development program on encryption, both of which have been very popular with waiting lists (see Table 2). Feedback for these sessions was (3.93/5) and (4.16/5) respectively. These sessions demonstrated that while staff may attend short sessions in RDM, they are more likely to attend session on tools. As we have so many researchers on our waiting list, we have pushed to get more resources available for self-taught training on our website with 'How To' guidance documents, online slides and notes, with additional training videos being made in the coming months.

The first attempt to coax researchers into attending our RDM session scheduled in the University's 'Generic Training for Researchers' (GTR) program in Feb. 2013, was postponed due to low levels of interest but a subsequent session in May 2013, went ahead with a modest attendance. We believed that the generic nature of this program does not attract researchers. This supports our DAF results where researchers said that they wanted targeted advice and that generic advice was too vague.

In response, we have focused on generic training with packages of subject slides with specific examples that can be inserted into the generic slides and delivered in research group training programs. We have also produced a 'Top Tips' presentation that can also be combined with the subject slides to be delivered in seminars. As researchers are accustomed to weekly seminars, the attendance should be higher than previous training sessions. Unfortunately, the seminar series are scheduled during the autumn and spring semesters so we cannot present these seminars until the 2013-2014 programs, which are beyond the life-time of this project.

We have successfully introduced the RDM training as a permanent topic in the CAR training program and into the bi-annual STRI welcome for new students. The encryption training will be carried out at least bi-annually depending on demand. Additional training session on the Document Management System (DMS) is planned for the next academic year, and bookings are being made for the seminar series throughout UH.

Table 2: Summary of the interest shown in training sessions throughout UH.

Date	Program	Title	Audience	Booked places	Waiting List	Attendance	Feedback average
16/10/12	Staff Dev.	RDM update	Staff	13	0	12	3.8
23/10/12	CAR	Intro. RDM	PGRS	6	0	4	4.3
04/12/12	Staff Dev.	Encryption	Staff	20	6	22	3.9
22/01/13	Staff Dev.	Encryption	Staff	19	3	14	4.2
30/04/13	Staff Dev.	RDM update	Staff	6	0	5	4.0
13/05/13	GTR	Intro. RDM	PGRS	19	0	14	3.8

3.2.5 Public training materials

The training courses described above cover the life-cycle of projects in four modules; planning a project, getting starting, safeguarding data, and the finishing touches. All of these modules include PowerPoint and pdf slides, notes for trainers, and videos for online self-taught learning. In addition, there are subject slides with subject specific examples and anecdotes, and activities to encourage the audience to put the generic training in perspective with respect to their own projects. We have published all of these resources on our RDM website (http://www.herts.ac.uk/rdm/training) with advice on selecting slides for reuse in other institutes and for a various audiences; researchers, PGRS, support staff, and technical staff. We have also linked to training materials produced by other JISC MRD training projects (DMTpsych, DATUM for Health, MANTRA, CAiRO, TraD and RDMRose) and reused their examples within our subject slides. We have also produced a slide selection table (see appendix) and advice on choosing the slides which will most benefit the audience given session time constraints.

All of these materials have been classified using the VITAE Researcher Development Framework (RDF) for use in their information literacy lens. These materials will be uploaded to JORUM with sufficient metadata for easy re-use by the community after final evaluation this autumn (see §3.2.6). At this stage, the material will be recorded into training videos using Camtasia for online, self-taught learning.

3.2.6 JISC program activity

Throughout the project we have conferred with other training projects and provided updates at MRD meetings (Table 3), which have described our progress from training for physics and astronomy, to physical sciences, and finally for the university with generic and subject specific training as well as our conclusions on the effective of the RDM tools for our researchers.

We have established links with the team at UEL working on the TraD project through these meetings and plan to utilise their training for support staff within UH in exchange for access to our TrueCrypt training materials. We have also participated in discussion with DaMSSI-ABC and representatives from the RDM training projects to establish connections between the JORUM and VITAE RDF archives and our training materials ensuring that they are compatible and increasing the likelihood of rediscovery.

JORUM is an archive of learning and teaching resources, shared by the UK Further and Higher Education community, and VITAE is supported by RCUK and UK HE funding to provide career development training for researchers. Both of these facilities provide affective sharing of these training resources; however, they are aimed at different audiences and include different keywords in their classifications. As key contributors of training materials, we have assisted in redefining the classification options so that RDM training materials are clearly identified in JORUM and so that the VITAE literacy lens can direct researchers to the relevant materials.

At UH, we have used the structure of JORUM to organise our training into topics and modules with metadata that can help others reuse our training materials and we have highlighted the primary and secondary learning outcomes achievable from each of the modules.

Table 3: Summary of the workshops and meetings attended by RDMTPA team members indicating the contribution to these sessions throughout the project.

Date	Meeting	Contribution
24 Oct 2012	Building Institutional RDM Services, Nottingham	Progress presentation (Dr Bill Worthington)
26 Oct 2012	RDM Training Stand workshop, London	Progress presentation (Dr Jo Goodger)
20 Nov 2012	DCC road show workshop at LSE, London	Participation in discussions (Dr Jo Goodger)
8 Mar 2013	Making citation work; practical issues for institutions (*DataCite*) at BL, London	Participation in discussions (Dr Jo Goodger)

25 Mar 2013	JISC MRD program workshop, Birmingham	Progress presentation (Dr Jo Goodger), poster presentation and participation.
18 May 2013	DaMSSI-ABC Training Strand workshop, London	Participation in discussions (Dr Jo Goodger)
14 June 2013	Research data metrics for impact and citation (*DataCite)* at BL, London	Participation in discussions (Dr Jo Goodger)
5 July 2013	National Astronomy Meeting (*Royal Astronomical Society)* at St Andrews, Scotland	Presentation "Preserving Digital Data at UH" (Dr Jo Goodger)

3.3 What did you learn?

RDM is not an attractive or engaging title for a training session
The sessions that we ran on 'Encryption with *TrueCrypt*' were fully booked; 36 attended out of 39 bookings (plus 9 researchers on the waiting list) compared to 17/19 researchers who attended our "Research Data Management Update" sessions. All of these sessions were run in the staff-development program and although the second session was more popular than the first session in both instances, the tool specific session was obviously more attractive to our researchers. The feedback suggests that running an "Introduction to RDM" session that takes only an hour is neither specific enough nor relevant enough to entice researchers.

We cannot force researchers to attend even the best titled training session
No matter how we rename the training sessions they will have one of the three words; Research, Data, and Management, which are going to put researchers off. Many researchers in arts, humanities, business, and social sciences do not consider their audio-video files, transcripts, photos, and survey responses as data. The term 'Management' conjures images of the business-related or bureaucratic aspects of a project, which is only of interest to those researchers who may be applying for large grants (eg. FP7 grants from the EU), or to researchers in the business school. An effective method may be to run a series of sessions named differently but with the same content to attract different researchers.

A simpler and more efficient method of engagement with training is to direct researchers to the online training materials and the guidance on our RDM website by using a series of posters warning of the risk caused by bad RDM and the benefits of good-practice.

Subject specific examples and anecdotes are key to giving the training context
We have seen that it can be difficult to attract engagement where training is not mandatory, or is generic compared with training badged as subject specific or tool specific. This is reflected in the responses to the DAF survey. To remedy this, we plan to target PGRS in their mandatory, subject training programs and to address researchers directly during their weekly seminars. Both of these programs will be aimed at an audience within a single discipline so will include subject slides to demonstrate the relevance of the tools available to aid researchers at every stage of their project.

Targeting students in their first year leads to prolonged good-practice
The subject training programs are generally mandatory for first-year students and as such give us the opportunity to teach best-practice RDM to researchers at the very beginning of their careers. Within 3 years, all of the PGRS will have had RDM training, and will therefore be prepared for a future where open data and RDM will be expected to be normal part of research practice.
The more difficult task is ensuring that their supervisory team, early-career researchers, principal investigators, fellows and professors all take on board the lessons of RDM. This may be achieved as and when DMPs become mandatory in grant applications and when penalties for non-conformance are evident.

RDM can be included in many existing training sessions
While a whole day session on RDM would be beneficial as it could incorporate detailed descriptions of the tools, processes, and considerations with examples for everyone, this would likely be too time-consuming for researchers as we have had comments that an hour is too long, and would require that the session be given by an expert in RDM with numerous examples and anecdotes to call upon depending on the attendees. A suitable

alternative to generic training might be to attach the topics to existing sessions. Within our staff-development program we already have successful sessions on; applying to research councils, project planning, and using the new Research Information System. Within our GTR we also have sessions on information gathering, research ethics, and project planning ("If you fail to plan, you are planning to fail"). All of these could benefit from a 10-minute add on of a relevant topic from our RDM training; however, as some are delivered by external experts, this may not be realised in these cases.

A website is crucial for 24-7 RDM availability
Having a comprehensive website of advice for each aspect of RDM that is openly available and openly assessable is vital to disseminating best-practice advice on RDM. Even if researchers are too busy to attend a training session, they can access the website at all times, on and off campus. The topics are organised based on the lifecycle of a project, but also alphabetically in our A-Z so that they are easy to find, and related advice within our RDM website, the UH StaffNet, the UH StudyNet, and the world-wide web are cross linked to give our researchers as many opportunities to find a solution to their RDM dilemmas as possible. The website also gives us a platform for our open, online material, which includes videos of the training slides on a topic-by-topic basis so that researchers can take 10-20 minutes to watch a video about a topic if they require more information than that which is included in the web pages.

Benefits are just as important as emphasising the risks and penalties
Throughout the training, we have put an emphasis on the benefits of using the tool, process, or best-practice recommendation to encourage researchers to adopt these methods. Many researchers had developed bespoke solutions for RDM such as producing their own website with object catalogues, whilst others limited the risk to their data by keeping it on paper and locked in a cabinet instead of sharing it electronically, even with collaborators. Our encryption training not only provided an open source tool which can be used on all operating systems to encrypt data on the fly, which can then be emailed and shared securely, but also gave our researchers the freedom to back-up and store their data securely. We discovered that a researcher in financial mathematics was already using *TrueCrypt* to encrypt his laptop, and that a researcher in robotics was previously limited to storing AV data of his robot on a single drive, sharing with collaborators fleetingly and only face-to-face. He is now using *TrueCrypt* to store his data electronically with a robust back-up policy which keeps his sensitive data secure.

Researchers are resistant to change without benefits
During our conversations with researchers, it became apparent that many of them were using *DropBox* to share data, including literature and documents, with their collaborators. As *DropBox* is subject to terms and conditions that conflict with the UH data policy we do not recommend using cloud storage at the present time. At UH we have networked storage which is secure, backed up, and accessible by staff and visiting members that we have championed to our researchers, but they are resistant as *DropBox* is very easy to access on all operating systems and does not require any special memberships or set up. Until researchers become aware of the risks of using cloud storage and the impact on them and their data, they cannot be convinced to use the UH solutions as they do not, to date, have a desktop folder and collaborators need to register with UH as visiting members.

A single point of contact would help
Whilst the RDM website offers many resources, signposts and pointers, as single point of contact such as rdm@herts.ac.uk is required for those researchers who prefer to ask than self learn.

3.4 Immediate Impact

Our work on the DMP has directed impacted the UH Data Policy and completion of the UH DMP at DMP*online* is now required for all of the projects at UH including post-award. This, combined with the funding body requirements, has made researchers aware that a change in how research data is managed is imminent.

Bringing together training
Researchers have benefited from the tools collated in the RDM Toolkit project and training on RDM is being rolled out across UH at all levels of research. During the summer, final adjustments will be made to the

training schedules of department training programs and to seminar series, all of which will include RDM. **All the known training channels throughout the university are aware of RDM training and beginning to use it.**

More secure data

We have already seen an increase in the number of researchers encrypting their data using *TrueCrypt*, and have received requests for encrypted devices. We have recommended *TrueCrypt* encrypted partitions instead of encrypted devices due to the risks of losing access to the entire device. These requests have demonstrated that researchers are considering the safety of their data more.

More reliable back-ups
Nearly every researcher that we spoke to had suffered from data loss or experienced a near-miss when a drive has failed and a back-up was either old or damaged. The damage and/or loss of data cost researchers both time and money. Our training has includes strategies for backup, including better use of networked storage, and we have seen an increase in safer practice as a result.

Greater awareness of risks, benefits, and responsibilities
During this project, research groups in astronomy, history, and clinical trials have tested our RDM tools and have consequently received more training than other researchers in the University. There has been a marked increase in requests for storage, DMP advice, and encryption advice from these groups as a result. This shows that **the level of RDM awareness among these groups has increased.**

3.5 Future Impact

Change in culture of research
As training is rolled out throughout the programs at UH, we will see a change in the methods used at all levels of research. Improved RDM by early career researchers will be carry with them into their careers and therefore propagate a change in the culture of research. Our seminars and staff-development training, combined with training early career researchers will lead to better RDM throughout the University and in turn enable us to bid for and plan future projects more effectively.

A website for the future
The RDM website will be incorporated into the research pages of the University's main website, while the RDM team will maintain content to keep it fresh and up to date with the changes in national policies and the further develop of institutional archives and subject repositories. Our website is complete with Google analytics which will monitor the popularity of topics, highlighting those which are most in demand and could therefore form training sessions in the future.

Our modular approach, in which we have produced generic guidance for UH staff with additional subject slides for the physical sciences, could easily be expanded upon to include more examples in other disciplines, thus extending the each of our materials. They could also be used as a template for other subjects that are not studied at UH.

Global Astronomer and Physicist awareness
We have also consulted with 31 Astronomers and Physicists from other institutes to determine the position in other research groups around the world. We expected differences between the institutes, influenced by their local infrastructures, and so were surprised to find that astronomers have similar methods and issues across the world. The funding process is universal and consequently, the plans for preservation of data are also similar. As a consequence we believe our materials, even those quite specific to UH, will be of use throughout the discipline. These Astronomers also highlighted that the infrastructure in CAR at UH is similar but often better than at other institutes; few have dedicated networks and often they have to find their own storage for their data, which can be up to a dozen TB of data during a computation intensive, modelling project.

We talked to some physicists who have recently left academia and still find that their employers deal with data in a familiar way. They deal with sensitive information when dealing with commercial partners but have shared data willingly on a personal level in the past when asked by interested parties with a genuine interest. However, many agree that archiving and preserving data in an 'open' context would benefit the physical sciences and would share their data in this way in the future.

We will be attending a session in July 2013 at the National Astronomy Meeting in St Andrews, which is an annual meeting of astronomers supported by the Royal Astronomical Society. Our work on preserving research data will be presented during a discussion on 'Preserving Astronomy Heritage'. The session is mainly to discuss how, when and what historical records should be preserved and produce an article of guidance. Our experience with preserving research data and the preservation of digital data from our Bayfordbury Observatory should be of interest to those in attendance and aid in the development of a preservation policy.

4 Conclusions

Through interviewing researchers, this project has determined that within UH:

1. Astronomers, and researchers in STEM subjects, have similar needs to those in other research areas at UH; the main differences are the quantity of data that they use, and the funding and preservation procedures, not the requirements.
2. Students and staff should have equal access to tools during their research projects; this can be achieved at UH by allocating visitor membership to students.
3. Many tools are not immediately accessible to researchers in the physical sciences as they work on Linux and Mac operating systems; we have suggested they use Windows Virtual Machines that are set up so that the network connections are not changed.

With respect to training specifically, we have found that:

4. We cannot force researchers to attend; web-based advice may be better as it is available 24-7 and topic add-ons included in existing training spread the word and make good RDM part of everyday research.
5. Seminars and welcome sessions are already regular parts of the academic calendar and should be used to spread best-practice RDM to researchers. These should include subject specific advice, which is key to demonstrating the benefits of best-practice RDM.
6. Referring to a session with 'Research Data Management' within the title is vague, uninspiring and unattractive to researchers. It suggests a generic course, is not clear on the contents, and was not well attended. It is better to give training on topics or tools, with discipline specific examples.

Consultation with the wider research community revealed that:

7. Within any research community, practices are generally the same irrespective of the institute.
8. Being able to share data, including documents, with collaborators outside of the university is crucial; an intra-institutional solution is not sufficient for global research communities.
9. Many researchers like the concept of preserving and reuse, but are afraid of abuse by media, competitors, and consequent loss of funding, opportunities, and credit.
10. Subject repositories may exist, but are often insufficient for community needs; what should be preserved is a complicated issue, but the guidelines and requirements need to be uniform throughout research communities.

From this we have determined that JISC should be aware that:

11. Inter-institutional collaboration is vital to determining subject specific guidelines as institutions offer different local solutions that may not be available to all parties in a group. This also leads to establishing common methods and behaviour within a research community.
12. Global solutions, or even national ones, would benefit researchers and a united front will help to put these in place, for example; subject repositories, where identifying fields that would benefit from a data centre can only be done on large scales, and cloud storage, as terms and conditions can seriously risk data.

5 Recommendations

We recommend that at UH:

1. Visiting-staff status should be issued during enrolment to PGRS and to UG project students during the acceptance of project, valid for 1-3 years depending on length of project.
2. VWM be installed on all UH-networked Linux and Mac machines on the staff network.
3. RDM training is integrated with the training delivered in GTR, staff-development, and in departmental training programs ensuring multiple opportunities to gain RDM training.

4. Additional online videos of training produced for 24-7 access.
5. A 'Top Tips' seminar with an introduction to the benefits, risks, and penalties of RDM delivered as part of existing seminar series throughout the university. Subject slides would put RDM in content for the audience.
6. Split the training materials into topics to reduce the session duration and focus on a single issue or tool. The audience will know what to expect and are more likely to recognise the benefit of attending that session for example, 'Completing a Data Management Plan', which will soon be required during funding application.

Outside of the University, we suggest that:

7. The research community establish a national or global deal with a cloud provider to allow sharing of research data within communities in the UK.
8. Researchers are made aware of the risks to their data and legal ramifications of using cloud storage.
9. Clear guidelines are established with regards to which data should be preserved and clear definitions for metadata for rediscovery and accessibility of data by community and media.
10. Institutional repositories installed to fill in the gaps in data stored by current data centres and archives until the need for new subject archives are established.

6 Implications for the future

Participation within UH and remote use of our training by researchers in the wider community will demonstrate to them the advantages of good RDM.

Many of the topics covered in the training are not unique to UH and can therefore be reused by other institutions in other subjects and for other audiences.

These materials were originally written for researchers and PGRS, but can be presented to information services staff including or eliminating different levels of technical information depending on the service and technical experience of the audience. The training can therefore contribute to developing awareness of RDM beyond out walls.

As the area of research data management matures, many of the topics within these training modules will evolve and settle, and more topics may become relevant. There is a danger these materials may become obsolete without maintenance. The RDM team, which is being taken forward at UH (see final report for RDTK) will continued to update these materials, as well as the RDM website, to ensure a constant source of advice and assistance for researchers at UH.

Future iterations of the training materials will also be published on JORUM and VITAE RDF.

7 References

Research data management toolkit (RDTK), University of Hertfordshire, JISC funded 2011-2013
Postgraduate training for research data management in the psychological sciences (DMTpsych), University of York, JISC funded 2010 - 2011
DATUM for Health: Research data management training for health studies, Northumbria University, JISC funded 2010-2011
Research Data MANTRA, University of Edinburgh, JISC funded 2010-2011
Curating Artistic Research Output (CAiRO), University of Bristol, JISC funded 2010-2011
Training for Data Management at UEL (TraD), University of East London, JISC funded 2012-2013
RDMRose, University of Leeds, University of Sheffield, and University of York, JISC funded 2012-2013

8 Appendices

Included in these appendices are additional tables and figures that readers may find helpful in understanding the work described in the body of this report. We have included the results of the training session feedback questionnaire, the comparison of the UH DMP Template questions to those in other RCUK templates, the slide selection table that will assist trainers in selecting training slides from our materials, and the questionnaire sent to astronomers and physicists worldwide.

8.1 DMP Comparison

A full breakdown of the UH DMP template comparison to the RCUK templates as available in March 2013. The DMPs were numbered as follows; 1 – STFC, 2 – NERC, 3 – ESRC, 4 – MRC, 5 – BBSRC, 6 – AHRC, and EPSRC was N/A.

UH Generic DMP Template	1	2	3	4	5	6
1 Introduction and context						
1.2 Short description of the project's fundamental aims and purpose				✓	✓	
1.3.1. Funding body requirements relating to the creation of a data management plan	✓			✓		
1.3.2. Institutional or research group guidelines	✓		✓			
1.3.3. Other policy-related dependencies	✓				✓	
1.4.1. Date of creation of this plan	✓					
1.4.2. Aims and purpose of this plan	✓					
1.4.3. Target audience for this plan	✓					
8.2.3. Does this version of the DMP supersede an earlier plan?						
10.1 Contact details and expertise of nominated data managers / named individuals			✓			✓
10.2 Glossary of terms						
2 Data types, formats, standards and capture methods						
2.1 Give a short overview description of the data being generated or reused in this research.	✓	✓	✓	✓	✓	✓
2.1.1. Have you surveyed existing data, in your own institute and from third parties?		✓	✓		✓	
2.1.2. What existing data could you use or build upon?		✓	✓		✓	
2.2.3. Describe any access issues pertaining to the pertinent, existing data?		✓	✓		✓	
2.4.2. How will you manage integration between the data being gathered in the project and pre-existing data sources?		✓	✓			
2.3.2 How will you capture or create new data?			✓	✓	✓	✓
2.3.3 Which file formats will you use, and why?			✓	✓	✓	✓
What directory and file naming convention will be used?						
2.3.2. Are there any tools or software needed to create/process/visualise these data?						✓
Are there appropriate computing hardware, facilities, and resources to manage, store, and analyse these data?						
2.3.4. What criteria and/or procedures will you use for Quality Assurance/Management?			✓	✓		✓
2.5.1. Are the datasets which you will be capturing/creating self-explanatory, or understandable in isolation?		✓	✓		✓	✓
2.5.2. If not, what contextual details are needed to make the data you capture or collect meaningful?		✓	✓		✓	✓
2.5.3. How will you create or capture these metadata?		✓	✓		✓	✓

UH Generic DMP Template	1	2	3	4	5	6
2.5.4. What form will the metadata take?		✓			✓	✓
2.5.5. Why have you chosen particular standards and approaches for metadata and contextual documentation?		✓			✓	✓
3 Ethics and Intellectual Property						
3.1.1. Are there ethical and privacy issues that may prohibit sharing some or all of the data?	✓	✓	✓			
3.1.2. If so, how will they be resolved?	✓	✓	✓			
3.1.3. Is the data that you will be capturing/creating 'personal data' in terms of the Data Protection Act (1998) or equivalent legislation if outside the UK?	✓	✓				
3.1.4. What action have you taken to comply with your obligation under the Data Protection Act (1998) or equivalent legislation if outside the UK?	✓	✓				
3.2.1. Will the data be covered by copyright or the Database Right? If so, give details in 3.2.2.		✓	✓			
3.2.2. Who owns the copyright and other intellectual property?		✓	✓			
3.2.3. How will the database be licensed?		✓				
4 Access, Data Sharing, and re-use						
4.3.1. Which groups or organisations are likely to be interested in the data that you will create/capture?					✓	✓
4.3.2. How do you anticipate your new data being reused?					✓	✓
4.1.1. Are you under obligation or do you have plans to share all or part of the data you create/capture?	✓	✓		✓	✓	✓
4.1.2. If not, why will you not share your data?	✓	✓			✓	✓
4.1.3. If so, how will you make the data available?	✓	✓		✓	✓	✓
4.1.4. If so, when will you make the data available?	✓	✓				✓
4.1.5. What is the process for gaining access to the data?	✓	✓		✓	✓	✓
4.1.6. If you are under obligation (4.1.1), will access be chargeable? If so, please give details.		✓			✓	✓
4.2.1. Does the original data collector/creator/PI, retain the right to use the data before opening it up to wider use? If so, please give details.		✓	✓	✓	✓	✓
4.2.3. Are there any embargo periods for political/commercial/patent reasons? If so, please give details.	✓	✓	✓		✓	
5.3.2. How will you implement permissions, restrictions, and embargoes?			✓	✓		
5 Short-term storage and data management						
2.3.6. What is the ballpark size of the data being collected/created?		✓		✓		✓
5.1.1. Where physically will you store the data during the project's lifetime?		✓		✓		✓
5.1.2. What media will you use for primary storage during the project's lifetime?		✓				✓
What software will be used in storing and processing these data?						✓
5.2.1. How will you back-up the data during the project's lifetime?		✓	✓	✓		✓
5.2.2. How regularly will back-ups be made?		✓				✓
5.2.3. Who is responsible for back-ups?		✓				✓
5.3.1. How will you manage access restrictions and data security during the project's lifetime?		✓		✓		

UH Generic DMP Template	1	2	3	4	5	6
6 Deposit and Long-Term Preservation						
6.1. What is the long-term strategy for maintaining, curating, and archiving these data?			✓	✓	✓	✓
6.2.5. On what basis will data be selected for the long-term preservation?	✓			✓		
6.2.1. Will or should data be kept beyond the life of the project?						✓
6.2.2. If so, how long will or should these data be kept beyond the life of the project?						✓
6.2.3. If so, which archive/repository/central database/data centre have you identified as a place to deposit data?	✓			✓		✓
6.2.6. If the data includes sensitive data, how will you manage this over the longer term?		✓				✓
6.3.1. What metadata/documentation will be submitted alongside the datasets or created on deposit/transformation in order to make the data reuseable?	✓			✓		✓
6.3.2. How will this metadata/documentation be created, and by whom?	✓			✓		✓
6.3.3. Will you include links to published materials and/or outcomes? 6.3.4. If so, please give details.						✓
6.3.5. How will you address the issue of persistent citation?						✓
7 Resourcing						
7.1. Outline the staff/organisational roles and responsibilities for data management.		✓	✓			
7.2. How will data management activities be funded during the project's lifetime?						
7.3. How will longer-term data management activities be funded after the project ends?	✓					✓
8.1.1. How will adherence to this data management plan be checked or demonstrated?		✓				
8.1.2. Who will check this adherence?		✓				
8.2.1. When will this data management plan be reviewed?		✓				
8.2.2. Who will carry out reviews?		✓				✓
Summary						
Total questions in RCUK template	30	50	29	22	27	43
Number of questions in common with UH, which has 65 question	22	38	23	20	23	38
Percentage of UH template completed at funding stage	34	58	37	31	37	58
Themes:	**1**	**2**	**3**	**4**	**5**	**6**
Intro and context	✓		✓	✓	✓	
Existing data		✓		✓	✓	
New data			✓	✓	✓	✓
Metadata		✓	✓		✓	✓
Ethics and IP	✓	✓	✓			
Open access	✓	✓		✓	✓	✓
Embargo, patent, security	✓	✓	✓	✓	✓	
Storage		✓		✓		✓
Backups		✓	✓	✓		✓
Archiving		✓	✓	✓	✓	✓
Resourcing	✓	✓				✓
Themes in common	5	9	7	8	7	7

8.2 Feedback on training sessions

A summary of the feedback from training sessions throughout UH is shown below. The questions asked refer to a) the content, b) the length of session, c) the UH infrastructure, d) DMPonline, e) funder requirements, f) UH storage, g) sharing solutions, h) archiving, i) encryption, j) the document management system (DMS), and k) the overall experience. Each was scored out of 5 where 1 strongly disagreed and 5 strongly agree with the statement.

Date	Program	Title	Aud.	Feedback											
				a	b	c	d	e	f	g	h	i	j	k	av.
16/10/12	Staff Dev.	RDM update	Staff	4.3	3.6	3.7	4.4	3.8	3.7	3.7	3.8	3.3	3.9	7.0	3.8
23/10/12	CAR	Intro. RDM	PGRS	4.3	4.3	4.3	4.3	4.5	4.5	3.8	4.0	4.0	4.5	4.5	4.3
04/12/12	Staff Dev.	Encryption	Staff	3.9	4.0	3.7						4.0	4.1	4.1	3.9
22/01/13	Staff Dev.	Encryption	Staff	3.9	4.3	4.1						4.1	4.3	4.3	4.2
30/04/13	Staff Dev.	RDM update	Staff	4.0	4.0	3.4	4.2	4.2	4.0	4.0	4.0	4.0	4.4	4.0	4.0
13/05/13	GTR	Intro. RDM	PGRS	3.9	4.1	3.6	4.5	3.8	3.6	3.6	3.0	4.3	4.0	3.9	3.8

8.3 Slide selection for training

The topics and activities for each audience can be selected using the following table as a guide. The duration of the session should also be taken into consideration.

	Topic length v		Researchers			PG Research Students			Services Support Staff			Technical Staff			Gen	UH
	Session Duration >	mins	1 hr*	½ day	1 day	1 hr	½ day	1 day	1 hr	½ day	1 day	1 hr	½ day	1 day		
PLANNING	Introduction to RDM	5	■	■	■	■	■	■	■	■	■	■	■	■		
	RDM Benefits and Penalties	5	■	■	■	■	■	■	■	■	■	■	■	■		
	DMPs	5	■	■	■				■		■					
	Project Lifecycle	20	■		■											
GETTING STARTED	Get Going Introduction	5			■	■										
	Filing Systems	5			■		■	■				■	■	■		
	Metadata: Data about your data	10		■	■		■	■			■	■	■	■		
	Software	5			■			■		■	■	■	■	■		
	Documents	5			■	■	■	■		■	■	■	■	■		
	Best Coding Practice	10			■	■	■	■				■	■	■		
SAFEGUARDING DATA	Storage	10			■			■	■	■	■		■	■		■
	Keep it safe – back up!	5		■	■	■	■	■	■	■	■	■	■	■		
	Remote Access to UH	5		■	■		■	■	■	■	■		■	■		■
	Sharing	10			■			■	■	■	■		■	■		■
	Security	10		■	■		■	■	■	■	■	■	■	■		
FINISHING TOUCHES	Publishing	10			■	■	■	■		■	■			■		
	Preserving	10		■	■	■	■	■		■	■		■	■		
	Repositories	10			■			■	■	■	■		■	■		■
	Web Publishing	10			■				■	■	■			■		
ACTIVITIES	Why should you share data?	10	■	■	■											
	What your data goes through?	10	■	■	■		■	■			■			■		
	What is data?	10		■	■	■	■	■	■	■	■	■	■	■		
	What risks affect your data?	10		■	■	■	■	■		■	■	■	■	■		
	What happens now you're published?	10		■	■			■		■	■			■		
EG.	Subject examples	1 ea.	■	■	■	■	■	■								

8.4 Global survey of astronomy and physics researchers

We contacted astronomy and physics researchers worldwide, including 13 across the UK, to determine whether the issues that affect the researchers at UH are also concerns for researchers globally. The map below shows the distribution of the researchers that were contacted, red cones are astronomers, and orange cones are physicists. (19 responded by 19th June – those shown with darker outlines).

And a closer look at the UK responses:

We asked our contacts 10 questions covering their project lifecycles:

1 - Who funds your project?
2 - What sort of data do you deal with now?
3 - Where do you get your data from?
4 - Where do you work on it i.e. laptop, desktop, networked drive?
5 - Do you back-up your work yourself or does IT do it?
6 - How do you back up?
7 - Where do you back up to? i.e. external HD, server, off-site PC?
8 - Do you deal with sensitive data? Do you need to encrypt anything?
9 - Do you publish the data when you publish the paper?
10 - How do you feel about open access to data after publication?