

THE UNIVERSITY OF LEEDS
Vice Chancellor's Executive Group
Funding for Research Data Management: Interim

SOME CONTENT HAS BEEN REMOVED FROM THIS PAPER TO MAKE IT SUITABLE FOR PUBLIC DISSEMINATION

VCEG is asked to consider this request for interim funding of £148k to enable work on the development of a research data repository and supporting infrastructure (Research Data Leeds) to be taken forward. This will ensure that the University is able to meet the funding councils' requirements in this area by May 2015 and provide the tools and support to enable academic staff to more effectively manage and share the data arising from their research. One of the key activities in this phase of work will be to develop a long term sustainability plan for the service.

Background

1. The need to manage, store and curate data arising from research is both an institutional and funding councils' requirement. In April 2011, EPSRC wrote to all university Vice Chancellors setting out a framework covering access to, and management of, research data arising from the research it sponsors. Their policy is that by May 2015 all organisations in receipt of grants will have put in place institutional arrangements to meet its data management framework requirements¹ and this has become a de facto target for all of the research councils.
2. The Research Councils and Wellcome require that data from projects they fund must be managed and made available long term. At a recent workshop on funding for the management of research data, the funders re-iterated the RCUK Common Principles on Data Policy.² They highlighted, in particular Principle 2 that says:

'Data with acknowledged long-term value should be preserved and remain accessible and usable for future research'

Existing Policies

3. A *Research Data Management Policy* was adopted by Senate in May 2012 (Senate Paper S/11/40). The aim of adopting an institutional policy and improving local infrastructure, training and support is to ensure that all research data assets are managed in a way that both meets the requirements of our funders and is appropriate to a world class research intensive university. The full text of the *Research Data Management Policy*, which lists a range of implementation benefits, is included here as Appendix 1.

¹ EPSRC Policy Framework on Research Data

<http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>

² RCUK Common Principles on Data Policy <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

In brief the benefits include:

- support for the re-use of data to enable new areas of research
- improved data integrity, security and access management, saving valuable time
- opportunities for further research collaboration
- improved research reproducibility and validation
- data is available for future generations
- further development of research skills
- improved citation rates from research outputs that also publish underlying data³
- the ability to cite data as a publication, which is likely to be a particular focus for REF-2020
- improved research competitiveness and research reputation
- improved relationship with research funders

Components of a research data management support service

4. Research data management requires effective systems *and* support services which will require different service areas to work closely with researchers in the faculties. The diagram below, developed by the Digital Curation Centre, aims to visualise the key service components identified under Jisc’s two major Research Data Management programmes (2009-2011; 2011-2013).

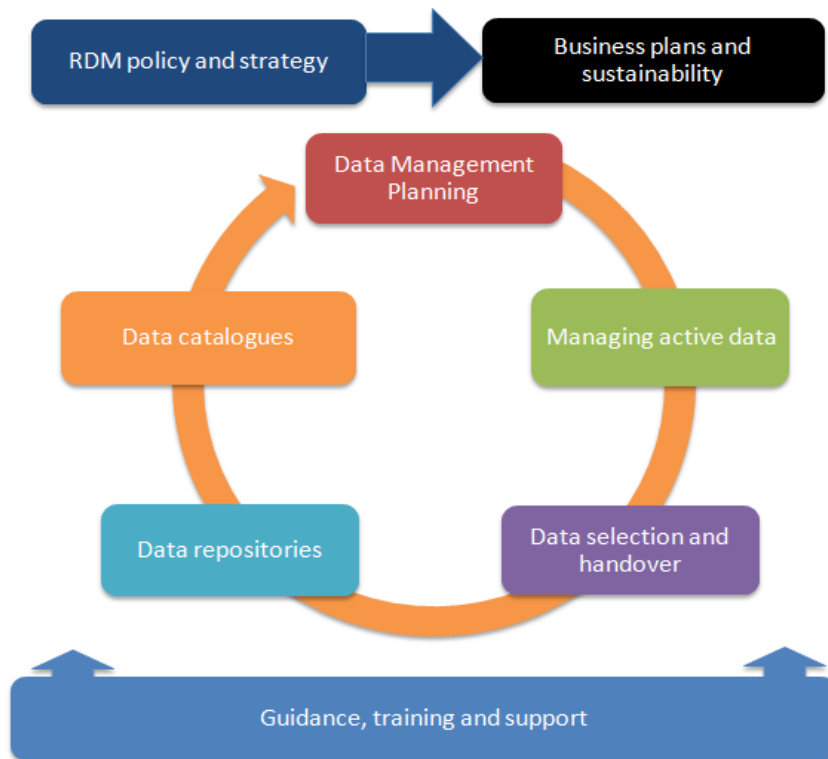


Figure 1: Components of RDM support services (Digital Curation Centre)

³ Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. [doi:10.1371/journal.pone.0000308](https://doi.org/10.1371/journal.pone.0000308)

Implementation at the University of Leeds

5. Implementation of the Policy and the development of supporting infrastructure has been led by the University's Research Data Steering (RDSG) and Research Data Working Groups (RDWG) established in June 2011, informed by the work of the Jisc funded 18 month RoaDMaP project (Research Data Management Pilot) which commenced January 2012. The work has been split into a number of work packages, each of which will move through the following four phases:

Phase 1: Scoping

During this phase, the high level requirements for each work package will be agreed and an 'as-is' assessment of the current practices at Leeds will be conducted. The resulting gap analysis will determine the amount of effort required and the timescales in which this can be delivered.

Phase 2: Planning

The planning phase involves drawing up a plan for delivery based on the results of the gap analysis and further more detailed work on the requirements for each element of the service. During this phase, the team will work closely with the OneIT Enterprise Architecture Team to ensure any new systems are fully integrated with the existing technology landscape. The work will also include mapping out key processes and reporting requirements as well as carrying out a series of small scale pilot activities.

Phase 3: Implementation

During the implementation phase, the various service elements will be moved from a detailed design to an operational status. There will be additional activities during this phase to promote the service elements and provide additional training / support for their use. Part of the work towards the end of this phase will be to move the services from 'project' to 'business as usual', supported by a sustainable funding model.

Phase 4: Operation

This is the post project phase during which the service will be supported and managed as business as usual

The key activities within the different work packages through all four phases are laid out in Table 1.

6. The team will work with key stakeholders (e.g. individual researchers and research groups, representatives from Faculty Research Teams, RIS and research funders, staff from the library, OneIT and SDDU) during all phases to ensure requirements are fully understood.

Table 1: *Research data management - phased development and implementation*

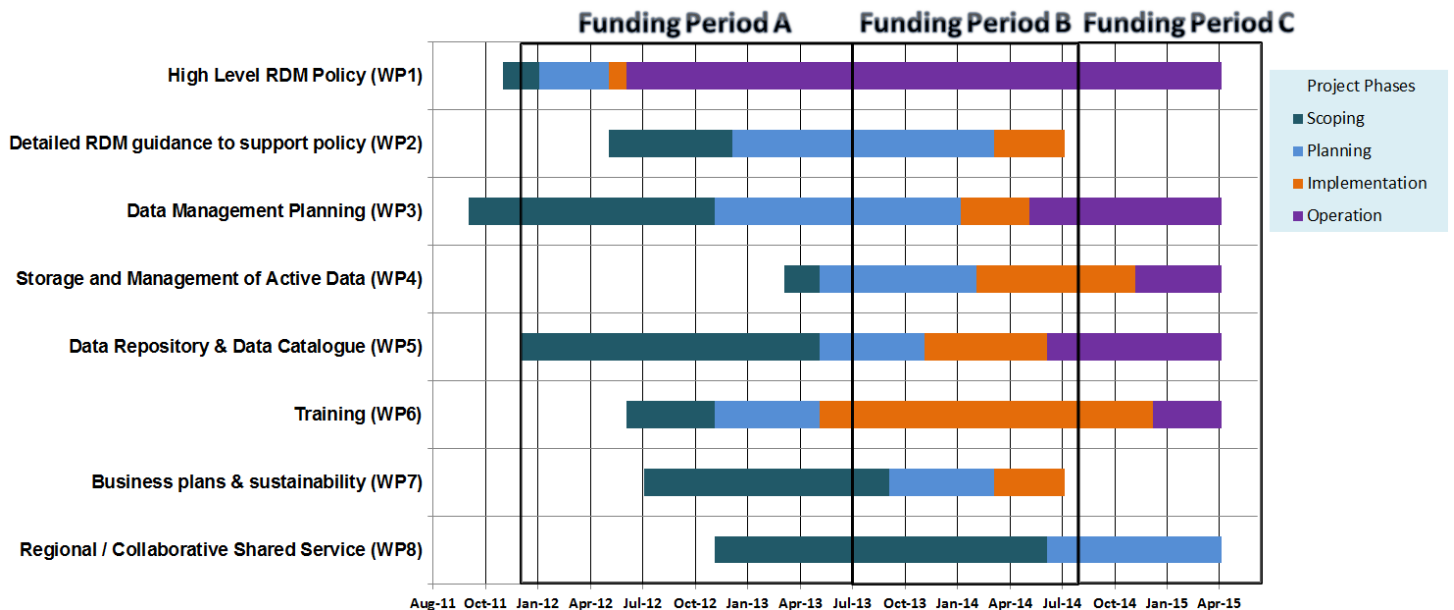
Phase	Activity
Phase 1 (Scoping)	<ul style="list-style-type: none"> ✓ All Establish governance structure (Research Data Steering and Working Groups) ✓ WP1 RDM Policy written and approved; EPSRC Roadmap submitted ✓ WP3 Pilot project with online Data Management Planning tool ✓ WP4 Pilot project with F5 networks for storage of active data ✓ WP5 Development of functional requirements for data repository; testing carried out with 3 software platforms; ✓ WP6 Pilot training sessions delivered to researchers, support staff; options paper for future training activities ✓ WP4 University-wide survey of researchers to better understand scale and nature of &5 research data being generated ✓ WP7 Initial scoping work on long term funding options ✓ All Work with three research project case studies ⇒ WP8 Investigate options for shared service delivery of some elements (e.g. WR / N8)
Phase 2 (Planning)	<ul style="list-style-type: none"> ⇒ WP1 Promote the policy amongst researchers, PhD students and support staff ⇒ WP2 Expand guidance and supporting website ⇒ WP3 Work with DCC to implement remaining changes to online data management planning tool; Embed DMP in research workflows; Link with KRISTAL ⇒ WP4 Define requirements and sustainability model for active data storage system & start procurement activity ⇒ WP5 Finalise detailed requirements for repository and investigate delivery options ⇒ WP5 Develop metadata schemas & data ingest mechanisms ⇒ WP5 Integrate with other system through Enterprise Architecture principles ✓ WP6 Start to deliver training to researchers, PhD students and support staff; ○ WP6 Develop online training materials ⇒ WP7 Develop business case with options for long term sustainability ○ WP8 Take forward discussions on shared service options with WR / N8 etc.
Phase 3 (Implementation)	<ul style="list-style-type: none"> ○ WP1 Continue to increase awareness of RDM policy ○ WP2 Fully roll out support / advice service ○ WP3 Implement data management plan requirements for every research project application / award ○ WP4 Put funding model for active data storage into full operation; increase capacity of active data storage facility ○ WP5 Move pilot repository through to live service and link to archive storage ○ WP6 Fully establish training programme; promote online training resources ○ WP7 Maximise recovery of RDM costs from research awards
Phase 4 (Operation)	<ul style="list-style-type: none"> ○ WP1 Policy widely published and understood by research community ○ WP7 Research data management activities becoming part of normal research practice ○ WP7 Cost recovery for data management activities maximised ○ All Data generated by UoL staff being used by other researchers & vice-versa

✓ Complete

⇒ **Started, but not yet complete**

○ **Not yet started**

The timing of the eight work packages is laid out in the following Gantt chart.



Funding

7. The funding required to implement the elements of good research data management across the University can be split into four periods, as follows:

Funding Period A (1st January 2012 to 30th June 2013)

8. This period was funded by a £250k grant from Jisc through the RoaDMaP project supplemented by contributions in-kind from the University, mainly staff time from a range of individuals who were involved in the project plus inputs from external project partners that came to £300k over 18 months.

Key deliverable are as follows:

- University RDM policy drafted and approved
- Scoping activities completed for all work packages, apart from the shared service work (WP8) where the rate of progress is determined by external factors
- Much of the planning work has been started and in some cases (e.g. training, WP6) completed
- Strong links have been built with other institutions both regionally and nationally through a series of events, workshops, meetings and informal discussions

Funding Period B (1st July 2013 to 31th July 2014)

9. As the Jisc funded project finishes at the end of June 2013, there is a need for a period of interim funding to 2014 to enable the work on the specification and development of a research data service to be taken forward and for the University to meet the RCUK funders' deadline for an operational service with appropriate supporting infrastructure by May 2015.

Key deliverables of this period will be as follows:

- A library-based research data support and advisory service
- First release of a University EPrints⁴ based research data repository available for use
- Complete plans in place from all work packages (with the exception of shared services work for the reasons outlined above)
- Data Management Planning online tool available for use by researchers (developed by the Digital Curation Centre)
- Approved business case for full service, including sustainability plan

10. The cost to deliver the above is c. £250k. However, some of the posts / staff time is being funded through core budgets (e.g. grade 7 developer posts in library systems team; time from staff in the library, ISS / OneIT, RIS, SDDU & Finance as well as researchers), so the contribution requested from strategic funds is £148k. This amount would continue to fund the following posts that have been in place throughout the RoaDMaP project:

- Research Data Interim Coordinator (0.6FTE Grade 8)
- Research Data Officer (0.5FTE Grade 7)
- Data Management Planning / Research Data Storage Lead (0.4FTE Grade 8)
- Research Data Repository Lead (0.5FTE Grade 8)

Supporting researchers with the management, archiving and curation of their data requires a multi-disciplinary team approach. Full details of the tasks to be undertaken by each of these roles during the interim funding period are included in Appendix 2.

11. In addition, the funding would cover a number of non-staff costs, as follows:

- **Archive storage pilot:** One area of considerable interest across the sector at the moment is long-term data archive-as-a-service options. One of the main benefits is that the POSF (pay once store forever) cost model allows the costs of long term archive data storage to be added to many research grants. This position was confirmed by RCUK in a recent meeting on funding for research data management. During this period, it is proposed that a full & detailed investigation is carried out into this data storage model with a number of large data sets. The costs of purchasing the hardware necessary to run this service, and for storing 30,000GB of research data for an appropriate period, is c. £30k. If the pilot is successful, the potential savings to the University are very significant. The same hardware could be used for the production service.
- **DataCite Subscription:** a British Library service that assigns DOIs (Digital Object Identifiers) which will be required for all data sets funded through RCUK awards. The ability to assign DOIs will make it easier for researchers to cite their data sources and associate them with their publications so helping compliance with RCUK funder requirements.
- **Development of online training materials for researchers:** One of the outcomes of the RoaDMaP project is a proposal for the creation of a tutorial based on the format used by the

⁴ <http://www.eprints.org/>

soon to be launched *Information Security Awareness Training*⁵. This tutorial cost approximately £30k to produce but these costs were shared as it was developed as a collaborative activity with several other institutions, who we believe would want to collaborate on a new online training resource in the area of research data management. In the expectation that costs would be shared we are proposing a budget of £8k to cover the Leeds contribution to the development of this resource.

- **Travel and attendance at RDM meetings:** A small budget for travel to enable members of the team to attend meetings and workshops of which there are a number planned over the next 12 months. These have been extremely useful in helping to shape our thinking through the sharing of knowledge and experience between us and other peer institutions.
- **Awareness raising and communications:** It will be crucial to raise awareness of the University's RDM policy and options for RDM planning and support. It is envisaged several awareness raising events will be organised and supporting printed material produced.

Proposed Budget for Funding Period B: 1st July 2013 to 31st July 2014

Breakdown of Costs £ inc. VAT (where applicable)		2012/13	2013/14	Total
Staff costs				
60% Grade 8		2,562	30,738	33,300
50% Grade 7		1,515	19,762	19,762
40% Grade 8			24,277	24,277
50% Grade 8			24,923	24,923
Total Staff Costs			99,699	103,200
Non staff costs				
Archive storage		0	30,394	30,394
Datacite Subscription		0	1,500	1,500
Development of online training materials for researchers		0	8,000	8,000
Travel and attendance at RDM meetings		0	2,000	2,000
Awareness raising and communications		0	6,000	6,000
Total non-staff costs		0	48,894	48,894

⁵ <http://iss.leeds.ac.uk/info/357/isms/>

TOTAL		4,077(1)	147,594	151,671
--------------	--	-----------------	----------------	----------------

- (1) The staff costs for the period 1st – 31st July 2013 have been covered from a small underspend in the Library budget for 2013
(2) The two Faculty-based secondments would be predicated on funding to provide backfilling.

Funding Period C (1st August 2014 onwards)

12. When the various service elements are ready to go-live, initial effort (over and above that required to deliver the business as usual service) will be required to deliver training, raise awareness, ingest data into the repository and help researchers develop new skills in this important area. The precise size and scope of this task will be defined during the interim funding period, but it is likely to be similar to that required to launch the VLE.
13. Some initial proposals for long term funding of RDM activity have been developed with help from Finance but work on them had been on hold while the funding councils developed their guidance on what could be charged. A recent meeting with funders provided some guidance but fuller information will be made available in the next couple of months. An initial paper for the RDWG included consideration of the following options:
- a. Inclusion of funding requests in bids (for allowable costs only)
 - b. Funding through the RAM
 - c. Using the HPC recharging model
 - d. Using another recharge model

It is likely that the final model will incorporate a number of the above elements. These and other options will be explored further during the planning phase of the work (WP7). Again, the size and scope of the effort required to support the service is likely to be similar to that required for the VLE service.

Shared Service Options

14. Initial work has been undertaken to consider what aspects of research data support it might be possible to offer on a shared service or collaborative basis. The White Rose Libraries are all leading initiatives within their institutions and are sharing their experience. Consideration of possible models for collaboration within White Rose is being undertaken by the Digital Curation Centre (DCC) as part of its Institutional Engagement Programme. The DCC is due to report on its work, which is not charged for, in August 2013. A paper drafted by the RoaDMaP team and the RDWG outlining possible options for collaboration on research data was discussed at a meeting of the N8 Library and IT Directors. Areas for possible collaboration or shared services were identified and four working groups were established to consider aspects of RDM provision. These Groups are to report back to a further meeting of the N8 group in the autumn.

Experience at other Universities

15. As one of the Jisc RDM projects, RoaDMaP has been able to keep in touch with developments in other Universities. Information gathered at recent meetings show that Leeds is in a similar position to most other research intensive universities in the development of services to support the

management of research data. Several institutions have started to put together teams to provide services. A brief summary is below:

- **University of Manchester:** 2.5 new posts (support / advice / training only); working to develop storage / repository / technical support solutions
- **University of Edinburgh:** Recent allocation of £2m to pump prime development of a service over 3 years
- **University of Bristol:** £2m invested in active data storage over last 5 years. Dedicated Research Data Storage Facility launched in 2011. 5 new posts (Service Director, 3x Research Data Librarians and 1x Technical Support post) to 'further explore' development of a RDM service.
- **University College London:** focus on storage for active data - £600k recently spent on storage for active research data. Initial team consists of; Head of Research Data Service; 1x Technical Architect; 2x support.

There are developments at Oxford and other institutions but most are in the same position as Leeds aiming to have services in operation during 2014 ahead of the EPSRC deadline of May 2015.

June 2013

Professor David Hogg, Brian Clifford, Rachel Proudfoot, Tim Banks

Appendix 1: The University Policy for the Management of Research Data

UNIVERSITY OF LEEDS

Research Data Management Policy

The management of Research Data reflects our:

- commitment to research excellence
 - recognition of our duty to our funders
 - appreciation of the value of our data - to us and to others
1. Research data will be managed to agreed standards throughout the research data lifecycle and according to funder requirements.
 2. Responsibility for research data management during any research project or programme lies with responsible owners such as Principal Investigators (PIs).
 3. The University is responsible for the provision of training, support and advice on research data management
 4. A data management plan that explicitly addresses the capture, management, integrity, confidentiality, preservation, sharing and publication of research data must be created for each proposed research project or funding application. Sufficient metadata shall also be created and stored to aid discovery and re-use. Data management plans should take account of and ensure compliance with relevant legislative frameworks which may limit public access to the data (for example, in the areas of data protection, intellectual property and human rights).
 5. All research data should be offered and assessed for deposit and preservation in an appropriate University, national or international data service or domain repository, unless specified otherwise in the data management plan.
 6. Data should not be deposited with any organisation that does not commit to its access and availability for re-use, unless this is a condition of the project funding or arising from other requirements.
 7. At the completion of each research project, the PI should ensure that all relevant research data are made available, subject to meeting appropriate requirements, in the location specified in the data management plan.
 8. Research and Innovation Board will be responsible for reviewing and updating the policy.

The University recognises the following benefits of implementing this policy:

- support for the re-use of data
- benefit future generations
- improved data integrity, security and access management
- opportunities for further research collaboration
- improved research reproducibility and validation
- further development of research skills
- the ability to cite data as a publication

- improved institutional research reputation
- improved relationship with research funders

Appendix 2: Staffing required for the interim period until 31st July 2014

Available on request. researchdata@leeds.ac.uk