Steps that HEIs can take to meet the EPSRC research data policy

6 March 2015

Authors
Sarah Jones, Digital Curation Centre (DCC)
Verena Weigert, Jisc.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>The significance of the EPSRC research data policy</td>
<td>4</td>
</tr>
<tr>
<td>EPSRC policy</td>
<td>5</td>
</tr>
<tr>
<td>Meeting the EPSRC’s expectations</td>
<td>7</td>
</tr>
<tr>
<td>About this guide</td>
<td>7</td>
</tr>
<tr>
<td>1. RDM policy, strategy, governance and sustainability</td>
<td>8</td>
</tr>
<tr>
<td>Policy development</td>
<td>8</td>
</tr>
<tr>
<td>Challenges</td>
<td>8</td>
</tr>
<tr>
<td>Sustainability and business planning</td>
<td>9</td>
</tr>
<tr>
<td>Challenges</td>
<td>9</td>
</tr>
<tr>
<td>2. Development of support services, increasing RDM capability and skills</td>
<td>11</td>
</tr>
<tr>
<td>Guidance and Training</td>
<td>11</td>
</tr>
<tr>
<td>Data management support services</td>
<td>12</td>
</tr>
<tr>
<td>Data management planning</td>
<td>11</td>
</tr>
<tr>
<td>3. Technical infrastructure and services for preservation and sharing</td>
<td>13</td>
</tr>
<tr>
<td>Active research data storage</td>
<td>13</td>
</tr>
<tr>
<td>Challenges</td>
<td>13</td>
</tr>
<tr>
<td>Data publication</td>
<td>14</td>
</tr>
<tr>
<td>Digital curation</td>
<td>15</td>
</tr>
<tr>
<td>Research data life cycle diagram</td>
<td>17</td>
</tr>
<tr>
<td>Conclusion</td>
<td>18</td>
</tr>
<tr>
<td>Bibliography</td>
<td>19</td>
</tr>
</tbody>
</table>
This guide suggests approaches to research data management (RDM) that will enable universities to meet the requirements of the Engineering and Physical Sciences Research Council’s (EPSRC’s) research data policy. We have included an overview of the EPSRC’s policy expectations, outlined particular implementation challenges and we offer some examples of how these have been addressed successfully. There are signposts to further resources and tools. This guide will be of interest to university senior managers and research data management support staff, including those who are responsible for giving advice to researchers on the storage, management, publication and archiving of research data.

The significance of the EPSRC research data policy

Research Councils UK (RCUK) has agreed seven common principles on data policy based on the common theme of access and reuse of research data and these principles underpin individual research councils’ policies, although the policies vary in their terms or specific requirements.

The EPSRC’s policy includes nine specific expectations concerning RDM, they assign primary responsibility for promotion of research data management to the research organisation and require it to provide systems, tools and support services to enable this. While the policy’s focus is on management of EPSRC-funded research data, it has had the wider effect of stimulating research organisations to think about RDM and the research data lifecycle in a way that goes beyond the requirements of a specific funder. This has inevitably thrown up a number of challenges for research organisations, especially with regard to resources. Many are concerned about how they can plan for these and sustain them in the long term.

The EPSRC policy framework was published in May 2011 and all research organisations in receipt of EPSRC funds were expected to develop a roadmap by May 2012, setting out how they would align their policies and processes with the EPSRC expectations. The Digital Curation Centre (DCC) offers a list of example roadmaps for review. The EPSRC expects full compliance with the expectations by 1 May 2015.

EPSRC policy

EPSRC has the following clear expectations of organisations in receipt of EPSRC research funding:

I. Research organisations will promote internal awareness of these principles and expectations and ensure that their researchers and research students have a general awareness of the regulatory environment and of the available exemptions which may be used, should the need arise, to justify the withholding of research data.

II. Published research papers should include a short statement describing how and on what terms any supporting research data may be accessed.

III. Each research organisation will have specific policies and associated processes to maintain effective internal awareness of their publicly-funded research data holdings and of requests by third parties to access such data. All of their researchers or research students funded by EPSRC will be required to comply with research organisation policies in this area or, in exceptional circumstances, to provide justification of why this is not possible.

IV. Publicly-funded research data that is not generated in digital format will be stored in a manner to facilitate it being shared in the event of a valid request for access to the data being received. This expectation could be satisfied by implementing a policy to convert and store such data in digital format in a timely manner.

V. Research organisations will ensure that appropriately structured metadata describing the research data they hold is published (normally within 12 months of the data being generated) and made freely accessible on the internet; in each case the metadata must be sufficient to allow others to understand what research data exists, why, when and how it was generated, and how to access it. Where the research data referred to in the metadata is a digital object it is expected that the metadata will include use of a robust digital object identifier (for example available through the DataCite organisation).

VI. Where access to the data is restricted the published metadata should also give the reason and summarise the conditions which must be satisfied for access to be granted. For example, ‘commercially confidential’ data, in which a business organisation has a legitimate interest, might be made available to others subject to a suitable legally enforceable non-disclosure agreement.

VII. Research organisations will ensure that EPSRC-funded research data is securely preserved for a minimum of 10 years from the date that any researcher ‘privileged access’ period expires or, if others have accessed the data, from last date on which access to the data was requested by a third party; all reasonable steps will be taken to ensure that publicly-funded data is not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK.

Summary of how the EPSRC policy affects Research Organisations

Research Organisations have primary responsibility for ensuring that researchers manage their data effectively. They need established infrastructure and processes to ensure:

- Retained EPSRC-funded research data is preserved for a minimum of 10 years
- Effective data curation is provided throughout the full data lifecycle
- Knowledge of publicly-funded research data holdings
- Discoverability, recording of third party access requests
- Notice and justification of access restrictions, for example: ‘commercially confidential’
- Awareness and use of relevant law, for example FOI
- Awareness and compliance with research data policies
- Adequate RDM resource allocation for example: from Quality-related Research (QR) funding or Research Grants

References:

2. Clarification and guidance on the interpretations of these expectations, epsrc.ac.uk/files/aboutus/
standard-clarifications/expectations-research-data-management/
Steps that HEIs can take to meet the EPSRC research data policy

Meeting the EPSRC's expectations

VIII. Research organisations will ensure that effective data curation is provided throughout the full data lifecycle, with ‘data curation’ and ‘data lifecycle’ being as defined by the Digital Curation Centre. The full range of responsibilities associated with data curation over the data lifecycle will be clearly allocated within the research organisation, and where research data is subject to restricted access the research organisation will implement and manage appropriate security controls; research organisations will particularly ensure that the quality assurance of their data curation processes is a specifically assigned responsibility.

IX. Research organisations will ensure adequate resources are provided to support the curation of publicly-funded research data; these resources will be allocated from within their existing public funding streams, whether received from Research Councils as direct or indirect support for specific projects or from higher education funding councils as block grants.

To monitor compliance after May 2015, the EPSRC will ask pro-VCs research to complete light touch ‘self-assessment’ questionnaires and will start with dipstick testing after the summer break by checking availability of data underpinning published research i.e. papers published on or after 1st May 2015. The EPSRC will investigate any complaints that an organisation it funds is failing to ensure research data is managed in line with their expectations. EPSRC aims to embed compliance checking through formal self-assessment and follow-up as part of regular dipstick visits by the Research Councils’ Audit and Assurance Services Group (AASG).

Failure to comply could lead to a loss of eligibility for EPSRC funding; however, this sanction is thought to be the worst case scenario, reserved for research organisations that demonstrably and unjustifiably obstruct the sharing of publicly funded research data.

Meeting the EPSRC’s expectations

To meet the EPSRC expectations, HEIs need to coordinate their stakeholders, departments and processes so that, together, they can deliver the diverse elements of a cross-institutional research data management service. Institutions must explore the national and international services that could complement or replace local solutions. For example, international identifier systems such as ORCID and DataCite could be integrated into existing local research management systems. The recently concluded Jisc-ARMA ORCID project saw eight universities work through institutional ORCID implementation and Jisc is now coordinating ORCID consortium membership for the UK. Institutions might also decide to make use of scholarly communications resources such as Thompson’s Data Citation Index to increase discovery of their data assets and demonstrate reuse.

The range of institutional policies and practices that could address the EPSRC’s expectations will be determined by a university’s research profile, its culture, mode of governance and organisational structure. In some research areas subject repositories are well established so the scale of operation within the university will be influenced by this. The Cambridge Crystallographic Data Centre, for example, is widely used for crystallography data and journals require authors to submit crystal structures to this database. But where suitable subject-based data centres are not available universities must ensure their research staff have access to effective systems and services for archiving and exposing the existence of their research data.

About this guide

For this guide, we have grouped the EPSRC requirements into three areas:

1. Overarching issues on research data management policy, strategy, governance and sustainability
2. Development of support services and increasing RDM capability and skills
3. Technical infrastructure and services required for storage, preservation and sharing

1. RDM policy, strategy, governance and sustainability

Policy development

The EPSRC expectations addressed here relate to institutional data management policies. They should exist for all data (Principle V), and universities need to engage with their researchers and students to ensure that they are aware of the EPSRC’s expectations and the regulatory environment that is relevant to research data as well as the way they can be shared (I). Where access is restricted the published metadata should give the reason and summarise the conditions which must be satisfied for access to be granted (VII). The full range of responsibilities associated with data curation over the data lifecycle will be clearly allocated within the research organisation and where research data is subject to restricted access the research organisation will implement and manage appropriate security controls (VIII).

Formal policies

To manage their research data effectively, many institutions need to develop a formal policy. A policy will help to clarify both what is expected by the different stakeholders within the institution and also who is responsible for each activity. Policies will inevitably reflect local organisational realities and so it is no surprise that the early adopters of RDM policies have taken very different approaches. The University of Hertfordshire, for example, has extended an existing institutional policy to include data management so that it covers all types of data (corporate, administrative and research) and it also includes an appendix giving more specific guidance on managing research data and the importance of RDM planning. In contrast the University of Southampton has augmented its policy requirements with details of procedures such as responsibilities with regards to data management planning and links to further guidelines on issues like storage options. A list of existing university research data policies and a briefing paper describing how to develop such policies are both available from the DCC. Agreed guidelines, such as the UK Research Integrity Office’s Code of Practice for Research, outline common expectations for the collection, use, storage and retention of research data.

Challenges

Not all research data can (or should be) be made openly available. Ethical issues, commercial confidentiality and licensing requirements might all restrict access. Researchers need to know when such conditions apply (also Principles I&V) and universities must ensure that sensitive data, access requests and intellectual property rights (IPR) are correctly handled throughout the research lifecycle. Legal obligations under the Freedom of Information (FOI), IPR, Copyright, and Data Protection Acts must also be taken into account when developing a data policy. The DCC Guide ‘How to license research data’ explains the impact that licences have on future research and the potential pitfalls to avoid. Further resources include Jisc guidance on how researchers should respond when faced with an FOI request.

Managing Data Access

Guidelines for managing data access must also be in place to determine the conditions of use that third parties must agree to before they can access data. There may be restrictions on the redistribution of data to third parties but access to commercially sensitive data, for example, could still be granted subject to a suitable non-disclosure agreement. The Wellcome Trust provides useful advice on DMPs in its guidelines for researchers. This explains the value of anonymising and aggregating data, as well as the role of data access committees to manage access to more sensitive material.

Sustainability and business planning

The EPSRC expects that research organisations will allocate adequate resources from their existing public funding streams to support the management and curation of research data (IX). Research data are valuable assets created by universities and maintained within them. Carefully costed plans should be put in place to care for these assets in the long term. Capital and revenue budgets must be found and allocated, just as they would for any other core service.

Practical steps for HEIs

» Identify a senior management champion. Without high level support progress is often slow.
» Identify key pre-existing policies and their owners; try to secure them as allies, present RDM as a shared concern
» Ensure that the working groups responsible for implementation have representatives from all key departments: the library, research office, information technology/systems, legal/ethics, finance and – importantly – the research community. A senior academic, typically the PVC for research, should lead the top level working group.
» Ensure the RDM policy or guidelines address the issue of preservation, who decides what is worth keeping? How long for? And how?
» Do not underestimate how long it will take to adopt and implement a policy.

Challenges

For many universities the lack of certainty about what it will all cost is a major problem. Perhaps the most challenging issues are knowing how much storage to budget for and how readily (and how soon) researchers will deposit their datasets. These unknowns make it difficult to build up an evidence-based business case.

Identifying costs

As well as semi-variable storage costs there are initially some fixed costs to take on board, including the employment of staff to facilitate the RDM development process.

Jisc’s Keeping Research Data Safe projects found that staff are the major cost component overall and that there is a minimum base level of staff cover, skills and equipment required for any service; the Keeping Research Data Safe Factsheet is a useful resource. Initial outreach, acquisition and ingest have high costs and the cost of access to data through retrieval costs and bandwidth is also significant. Archival storage and preservation represent a very small proportion of the total. Ingest automation and leveraging shared services and economies of scale should help to reduce these costs for universities over time.

The UK Data Service’s costing tool and checklist helps institutions think about the resources they need when setting up a data service and RCUK funders have provided further clarifications on specific allowable RDM cost elements following a Research Data Management Forum event on funding RDM. Many universities have made an internal business case to secure core funding to establish and run their RDM services and are defining a basic level of support provision over and above which changes may be applied. The University of Bristol has published storage costs so that researchers can include any additional charges in their grant applications.

Useful examples of costing models for digital preservation and data curation can be found at the Open Preservation Foundation and we have included...
1. RDM policy, strategy, governance and sustainability

Examples of storage estimates and some default allocation quotas in the final section of this guide on active research data storage.

The Jisc co-ordinated 4C ‘Collaboration to Clarify the Costs of Curation’ project has developed the Curation Cost Exchange (CCEx) tool for digital curation staff to submit and compare costs of curation. The tool helps staff understand curation costs and also how they can use cost models to make suitable investments in curation. Organisations can also submit their curation costs to the tool to analyse and compare their own costs against those submitted by similar organisations, helping them to make informed decisions around curation investments.

Cost recovery

Universities are taking different approaches to recovering costs for storage and preservation of data. Some see this as core business and fund their research data management activities out of core funding. Others adopt cost recovery models that are triggered by individual data deposits. The University of Bristol uses a system modelled on the Princeton ‘Pay Once, Store Forever’ (POSF) model, which calculates a one-off charge for data storage that covers future refresh of the media by applying Moore’s Law to storage costs. The university has defined ‘forever’ in this context as 20 years. In contrast, The Bodleian Library at Oxford has established the University’s institutional data repository, ORA-Data, as a small research facility, to enable researchers to include local and remote or cloud solutions. Consider a consortium approach.

Work with researchers to include long term data management costs in grant applications.

Create policies on storage and preservation allocations, data quotas, monitoring use, and cost recovery.

Use the Jisc financial X-ray tool, which helps people cost new IT infrastructure choices and assists with moving to the cloud.

Practical steps for HEIs

- Engage with departments and researchers to understand the amount of data that needs to be preserved.
- Look for the most cost effective type of storage or preservation solution for the institution and model the costs between local and remote or cloud solutions.
- Consider a consortium approach.
- Work with researchers to include long term data management costs in grant applications.
- Create policies on storage and preservation allocations, data quotas, monitoring use, and cost recovery.
- Use the Jisc financial X-ray tool, which helps people cost new IT infrastructure choices and assists with moving to the cloud.

Guidance and Training

As pointed out above, research organisations are expected to promote the EPSRC’s research data management principles to researchers and staff (I). This is expected by the EPSRC to be an ongoing part of the normal business of organisation-based training.

Universities can meet this advocacy requirement by providing guidance and training. A wide variety of approaches have been taken which could be followed by others, for example:

- A number of universities have produced institutional RDM web pages, collating best practice and directing researchers to local support. These help to raise awareness of existing materials and provide a focus for support. A collection of such websites is available via the DCC.

The University of Northampton and University of East London have trialled data surgeries in conjunction with training events. This allowed researchers to ask specific questions related to their projects. Many universities are offering ad hoc one-to-one sessions, specifically to support researchers when writing data management plans (DMPs). This gives researchers the opportunity to get direct feedback and for institutions to review plans.

- Training in data sharing that is embedded within the academic curriculum for masters and doctoral students helps to engender data sharing practice early on. Openly licensed, reusable learning materials such as the research data MANTRA course can be adapted by other universities and embedded into their learning environments. RDM Rose provides an Open Educational Resource for information professionals on Research Data Management.

When developing training materials it is essential to collaborate closely with academic departments to create effective discipline-focused materials. Jisc’s Managing Research Data programmes created a number of discipline-specific training courses and these are available via Jorum. The EC-funded FOSTER portal is another useful place to locate training content on open science more generally.

Data management support services

Because data is part of the essential evidence base of a scholarly article, the EPSRC (and RCUK) principles on research data state that sufficient metadata should be recorded and made openly available. This enables other researchers to understand the potential for further research and reuse of the data. All published research papers should include a statement describing how and on what terms any underlying datasets may be accessed.

Data Citation

Just as other research that is used in a scholarly article should be cited, so should any datasets that are drawn upon to substantiate the published findings. Statements on how the underlying data can be accessed can take many forms, ranging from contact details to request forms or data management plans. The EPSRC recommends that each dataset has a unique identifier which can be cited. Citations must provide the reader with enough information to access the dataset. If the data cannot be made directly accessible the data citation should contain a persistent link to a separate document or landing page containing the associated metadata so that a future user can understand the potential value of the data and gain access under certain conditions if possible.
Examples of suitable citation formats are available from a variety of sources including DataCite and the DCC. National disciplinary data centres such as the UK Data Service provide advice on data citation for their sectors. The University of Bristol offers examples of data access statements that may be suitable for external datasets that might not be publicly available.

Data management planning

Policies and plans are expected to be in place for all existing projects (Principle V). Data management planning is the process of planning, describing and communicating the activities carried out during the research lifecycle in order to keep research data safe, maximise its reuse potential and to support its longer term preservation.

Researchers and institutions will benefit from support and services for planning activities especially before research data is created and collected. The DCC has worked closely with research funders and universities to produce DMPOnline, a tool that assists researchers to produce an effective DMP for the whole project lifecycle. It offers several templates representing the requirements of different funders (including one for the EPSRC) and universities. Users are asked a number of questions at the outset to determine the appropriate template to display and guidance is provided by research funders, universities and disciplines to help researchers interpret and answer the questions. Several universities have customised the tool to point to relevant institutional policies such as guidelines on research ethics and information about ethics committees on their RDM webpages. The EPSRC expects data management plans to be in place and the template can help researchers and institutions to address specific expectations.

Practical steps for HEIs

- Review existing guidance and support to identify gaps to be filled or potential to collate resources
- Provide RDM training, embedding this into existing provision and core/research skills courses where possible
- Reuse existing tools and resources to assist researchers with DMPs
- Provide clear guidelines to help researchers determine IPR and data ownership. This will have implications on data licensing and meeting funder requirements for openness
- Assist researchers with data licensing and citation to support the reuse of data
- Provide guidelines to help researchers determine which data to keep

Deciding what to keep

The EPSRC also expects research organisations to put policies and processes in place to identify their research data assets and the risks to these (III). It’s ultimately up to researchers to decide which data to keep but they may need support to make this judgement. What is judged relevant will vary by discipline. In the case of published research, what to keep should be informed by consideration of what people would need if they were seeking to validate or reproduce the results or to check the underpinning evidence used to support the findings. If the research data does not support published research findings then it’s again up to the researcher to decide what to keep, informed by their best assessment of the likely future value of the data. The DCC has a guide that addresses issues such as assigning responsibility for appraisal and how to create an appraisal and selection policy. It has also developed a useful checklist for appraising research data.

The EPSRC policy expectations being addressed here (IV, V & VII) relate to the appropriate storage of publicly-funded data and maintaining a record of holdings to facilitate access. Research organisations should ensure that data is not held in any jurisdiction where the legal safeguards provide lower levels of protection than are available in the UK. Moreover, non-digital data should be stored in a manner to facilitate it being shared in the event of a valid request for access. These requests for access should be tracked as this affects the period for which the data should be preserved. The EPSRC expects data to be securely preserved for a minimum of ten years from the end of embargo or, if others have accessed the data, from the last date on which access to the data was requested by a third party.

One way in which access requests could be monitored is via an institutional data catalogue. Research organisations are expected to publish metadata describing the research data they hold. The EPSRC recommends what should be included in this metadata and expects that it will be freely available online within 12 months of the data being generated.

Active research data storage

As EPSRC policy expectation VII requires effective data curation throughout the lifecycle, services must be made available during the data collection, capture, interpretation and analysis phases. This could be termed ‘active research data storage’ where data is work in progress and therefore subject to change; a high amount of collaboration and sharing is often needed.

Challenges

Given the high levels of use of third party cloud services such as Dropbox, Google Drive and Microsoft OneDrive amongst the research community, one of the key challenges to address is ensuring researchers have enough research data storage and flexible mechanisms for sharing with collaborators. Responses to the use of cloud services have been varied. Some institutions, such as the London School of Economics and Political Science (LSE), have provided pragmatic guidance on when it may be appropriate to use cloud services and for what kinds of data. Others have provided in-house alternatives by piloting open source tools such as OwnCloud or trialling the use of corporate solutions such as SharePoint for data. Jisc has also secured a deal for the sector to offer file sync and share services. This will allow institutions to purchase third party solutions at a preferential rate and gain some assurance that data is held in the European Economic Area (EEA) for data protection purposes.

Suitable storage options

It is important that institutions provide guidance to their researchers on suitable storage options as EPSRC policy expectation VII states that ‘all reasonable steps will be taken to ensure that publicly-funded data is not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK. This may rule out Dropbox and Google Drive as data is often held on US servers, but Microsoft OneDrive and Amazon Web Services, through Jisc-negotiated agreements, both offer storage on EU-based servers to institutions. The geographic location of data is even more important when it comes to sensitive material such as health data, which can be restricted to a specific part of the UK.

Storage requirements

In terms of storage, more work needs to be done by institutions to identify storage requirements for capacity planning. Several universities have carried out surveys to help determine the scale of current holdings and likely growth rates. The results of this work at the University of Hertfordshire led it to allocate 50GB per researcher as a default offer. However, the institution also established an RDM triage system with the PI for each new funded award and up to 5TB will be allocated on the basis of need. Other universities are providing differing quantities at the University of Edinburgh each researcher is allocated 5TB although this can be pooled by groups.
the University of Bristol offers 5TB per ‘data steward’ - a nominated data management contact, usually a PI, and the University of Manchester provides 20 TB per project, allocated only on completion of a compulsory DMP. Further examples are provided in the DCC’s briefing from its 2014 RDM survey.

Most universities provide details of charges so that any requirements above the quota can be costing into grant applications.

Non-digital data
The storage of paper records related to research has arguably been less well addressed. Some institutions have considered using corporate records management stores for data, while others are looking at routine digitisation of known classes of physical records such as lab notebooks. The EPSRC policy requires data to be stored in a way that allows retrieval and provision of access. Recording non-digital data in the institution’s catalogue and digitising it on request would be sufficient.

Data publication
When data becomes a ‘finished product’ and can be used to support research findings, it needs to be moved from active storage into a repository. Here, discovery metadata is created for the dataset, an identifier (such as a digital object identifier (DOI)) is assigned and the data (or metadata where access restrictions apply) is published. At this point the researcher and institution must select the appropriate publication space, which won’t always be the university repository. It might instead be one of the numerous national and subject-specific repositories or the data may have a publication location mandated by a journal publisher or research funder. A comprehensive list of data repositories can be found at re3data.org. Research Data Oxford has provided a data decision tree for EPSRC funded researchers in order to find a suitable data publication space for their data.

Metadata
The EPSRC has specific requirements around the provision of metadata for research data holdings. Sufficient metadata should be provided to allow others to understand what research data exists, why, when and how it was generated and also how to access it. A robust digital object identifier should also be assigned. There are several types of persistent identifier that could be used to identify datasets: examples include Handle, Archival Resource Keys (ARKs) and Persistent URLs (PURLs), all of which can be resolved to an internet location. Arguably the scheme that is gaining most traction is the Digital Object Identifier (DOI). The creation of a DataCite DOI necessitates a minimum amount of metadata such as the data creator, publisher, the publication year and the title of the dataset, and there are several other optional fields that can enhance the discoverability of the data through the DataCite search API. The additional identifier fields are useful to enable interoperability between other identifier systems such as ORCID for researchers and can be useful to create machine-readable links to publications, which can help to track the impact of research data.

Several universities have been developing data registers or catalogues. The University of Oxford had developed a standalone data discovery tool, DataFinder, as part of the Jisc-funded DaMaRO project, but in order to make their data services more user-friendly this has been replaced with an integrated data repository and catalogue service, ORA-Data. The new service offers a single point of deposit, discovery and access for users, and enables researchers to create rich metadata about their data to ensure it can be understood and reused. Work has also been done by the UK Data Service to define core metadata based on Dublin Core, DataCite, Data Documentation Initiative (DDI) and Inspire standards and this has led to the development of the ReCollect plugin for ePrints repositories. Other institutions are using their research information management systems to store the metadata. The University of Lancaster has decided to use Pure as its data catalogue, while Loughborough University is using Symplectic. Jisc, the DCC and the UK Data Archive will continue their collaboration to develop a Research Data Discovery Service. The first phase of this work trialled the feasibility of the Research Data Australia catalogue, harvesting data from nine universities and three data centres, developing a number of metadata crosswalks and engaging with stakeholders to understand their requirements. The second phase of work (November 2014 – July 2016) is exploring stakeholders’ requirements, piloting alternatives such as CKAN and laying firm foundations for a discovery service, including a service operation plan and a business case for its delivery into the future.

Digital curation
The EPSRC also expects that research organisations will ensure that effective data curation is provided throughout the full data lifecycle (VIII). This implies that the full range of responsibilities associated with data curation over the data lifecycle will be clearly allocated within the research organisation and where research data is subject to restricted access the research organisation will implement and manage appropriate security controls, research organisations will particularly ensure that the quality assurance of their data curation processes is a specifically assigned responsibility.
As the diagram opposite indicates, digital curation involves a number of activities that span a range of stakeholders. The creation and active management of data in the short term will be primarily a concern for researchers, though naturally they will draw on support and services from others. Digital preservation and any associated transformations such as migration will be the responsibility of institutional repositories and data centres who act as long-term stewards for the data. Other activities such as making content available are shared responsibilities that require significant input from service providers and the researchers who understand the data best. Clarifying roles and responsibilities and establishing sound workflows for the key points of interaction are key.

Managing retention
The EPSRC expectation that data will be preserved for ten years is similar to those of most other funders. What differentiates the implementation of the EPSRC requirement is determining the point at which the ten year period begins. This is either the point at which data first becomes available after any period of privileged access expires, or the last date on which access was requested by a third party. Records managers have not always been involved in RDM working groups, but their experience of monitoring access and managing retention periods will prove invaluable in dealing with the EPSRC requirements. Systems are needed to enable usage to be tracked and to alert the data service as to when data can be removed. Data may be reviewed to determine whether it still needs to be kept or if it should be disposed of as the ten year period elapses.

Data archiving
In terms of data archiving, the University of Edinburgh has undertaken a useful analysis of the functionality needed in a ‘vault’ service. Some institutions are operating their own data back-ups using offline tape storage and hierarchical storage mechanisms. Many others have chosen to outsource data archiving to Arkivum, an approved contractor with whom Jisc has signed a framework agreement. There are several factors to consider when outsourcing long-term storage, not least additional egress charges (costs attached to data retrieval and download) and any fair use policies that may be applied. It is important to try to model expected patterns of use and to understand the full cost of outsourcing before deciding which is the best option. Naturally, alternative homes may be available for a range of data via subject data centres and community initiatives. Most institutions recommend that researchers place their data in a subject repository if one is regularly used by their community, or the institutional repository if not.

> Assign responsibilities across services so it is clear who is responsible for each aspect of data curation. Projects should similarly assign roles across their teams for all activities involved in data management.
> Develop appraisal guidance to help determine which data should be kept.
> Provide services for long-term storage and sharing. These could be outsourced or delivered in partnership with others.
> Track the usage of the data to monitor impact and help with future curation decisions.
There are many aspects to address when establishing RDM support services so a phased approach to development is recommended. The challenges faced will vary by institution based on the existing infrastructure, research environment and prevailing organisational culture. Attention needs to be paid to the local context to determine the best starting point.

Securing the buy-in and support of others is critical. RDM requires the input of all university services; aim to convene a broad-based working group with the backing of senior research leads and the academic community. Input from a wide-ranging group should also help to ensure work is continued when staff leave or move on to new posts, and that the services developed meet the diverse needs of the research community.

Jisc is continuing to support the implementation of RDM solutions for UK universities through Research at Risk which aims to develop shared infrastructure, services and advice, and through Jisc’s support of the DCC which provides a number of useful resources, in particular the ‘How to develop RDM services’ guide. We are continuing to support universities with EPSRC compliance as part of Research at Risk, for example a workshop was run in February 2015 and a series of case studies focused on particular university approaches have been developed.

### Bibliography

1. RDM policy, strategy, governance and sustainability

#### Page 4
- RCUK Common Principles on Data Policy, rcul.ac.uk/research/datapolicy/
- EPSRC policy framework on Research data – Principles, epsrc.ac.uk/about/standards/researchdata/principles/
- EPSRC policy framework on Research data – Expectations, epsrc.ac.uk/about/standards/researchdata/expectations/
- List of example EPSRC roadmaps, Digital Curation Centre, dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps
- Summary how the EPSRC policy affects research organisations. Presentation slides by Ben Ryan EPSRC, prezi.com/kfylbtkcgyu/rdm-principles-and-expectations/ at Jisc/DCC workshop ‘meeting the EPSRC policy framework’ dcc.ac.uk/events/other-dcc-events/RDM-readiness on 13th February 2015

#### Page 5
- Clarifications of EPSRC expectations on research data management (2014), epsrc.ac.uk/files/aboutus/standards/clarificationsofexpectationsresearchdatamanagement/

#### Page 6
- Compliance monitoring of EPSRC research data policy after 1 May 2015. Presentations slides by Ben Ryan, EPSRC, prezi.com/kfylbtkcgyu/rdm-principles-and-expectations/ at Jisc/DCC workshop ‘meeting the EPSRC policy framework’ dcc.ac.uk/events/other-dcc-events/RDM-readiness on 13th February 2015,
- RCUK Responses to questions raised at the DCC/RDMF Special Event on funding for Research Data Management held at Aston University on 25th April 2013, rcul.ac.uk/files/2013/07/RCUK-Responses-to-DCC-RDMF-Funder-Questions-.pdf

#### Page 7
- Jisc - ARMA ORCID pilot project blog. orcidpilot.jiscinvolve.org/wp/blog/

Next steps for ORCID adoption: ORCID consortium membership for the UK (2016), orcidpilot.jiscinvolve.org/wp/2015/02/03/next-steps-for-orscid-adoption-orcid-consortium-membership-for-the-uk/
Steps that HEIs can take to meet the EPSRC research data policy

Bibliography

Page 8

UK Research Integrity office. Code of Practice for Research, ukrio.org/publications/code-of-practice-for-research/


Ball, A. (2012) How to License Research Data, DCC How-to Guides, Digital Curation Centre, dcc.ac.uk/resources/how-guides


Welcome Trust. Guidance for Researchers: Developing a data management and sharing plan, wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.html#four

Page 9


UK Data Service - Data management costing tool and checklist, data-archive.ac.uk/media/247429/costingtool.pdf

Archaeology Data Service costing calculator, archaeologyservice.ac.uk/easy/costing?sessionid=31b7b9d919cbeef0af2155ebeab6b

RDMF Special Event: Funding Research Data Management, a conversation with funders (2013), dcc.ac.uk/events/research-data-management-forum-rdmf/rdfm-special-event-funding-research-data-management

List of costing models for digital preservation and data curation, Open Preservation foundation, wiki.opf-labs.org/display/CDP/Home

Page 10
Curation Cost Exchange (CCEx) tool, curationexchange.org/

EU project: 4C Collaboration to Clarify the Costs of Curation, 4cproject.eu/

Policy for the use of Research Data Storage Facility, University of Bristol, acrc.bris.ac.uk/acrc/RDSF_policy_220814.pdf

2. Development of support services, increasing RDM capability and skills

Page 11
Collection of RMD Guidance webpages, Digital Curation Centre, dcc.ac.uk/resources/policy-and-legal/rdm-guidance-webpages/rdm-guidance-webpages

Jones, S (2012) The value of a one-to-one, dcc.ac.uk/blog/value-one-one

Mantra, Research Data Management Training, free online course, University of Edinburgh, datalib.edina.ac.uk/mantra/

RDMRose, Research Data Management Learning Materials, University of Sheffield, sheffield.ac.uk/is/research/projects/rdmrose

Jisc Managing Research Data Programme: Discipline Specific Training Courses, dspace.jorum.ac.uk/xmlui/handle/10949/18267

Training Content on open science on EC-funded Foster project portal, fosteropenscience.eu/

Page 12

Examples of Data Citation, DataCite, datacite.org/services/cite-your-data.html

UK Data Service, Citing data, ukdataservice.ac.uk/use-data/citing-data.aspx

Data access statements, University of Bristol, bris.ac.uk/expsych/research/brain/larg/research/collaborations/carta/data.html

DMPOnline, Digital Curation Centre, dmponline.dcc.ac.uk/
Acknowledgements

Thanks is due to Ben Ryan Engineering and Physical Sciences Research Council (EPSRC), Rachel Bruce, Jisc, John Shemilt, Imperial College & Russell Universities Group of IT Directors (RUGIT) for their useful comments on an early draft of this guide and to John Kaye, Jisc for his useful comments and contributions to Section I business planning and Section II.

The digital curation centre (dcc.ac.uk), provided by Jisc, provides expert advice, practical help and resources to anyone wanting to store, manage, protect and share digital research data in UK higher education and research.

Whyte, A. & Wilson, A (2010) How to Appraise and Select Research Data for Curation, DCC How-to Guides, Digital Curation Centre. dcc.ac.uk/resources/how-guides

DCC (2014) Five steps to decide what data to keep: a checklist for appraising research data v1, Digital Curation Centre. dcc.ac.uk/resources/how-guides

3. Technical infrastructure and services for preservation and sharing

Page 13
Using Dropbox and other cloud storage services, London School of Economics and Political Science. lse.ac.uk/intranet/LSEServices/IMT/guides/softwareGuides/other/usingDropboxCloudStorageServices.aspx

Jisc Service, File Sync and Share, ja.net/products-services/janet-cloud-services/file-sync-and-share

Rans, J (2013) Defining institutional data storage requirements, dcc.ac.uk/blog/defining-institutional-data-storage-requirements

Page 14
Archiving and citing research data to meet EPSRC’s requirements, Research Data Oxford, researchdata.ox.ac.uk/epsrc-data-requirements-and-what-you-need-to-do/

Page 15
Data finder service, University of Oxford, blogs.it.ox.ac.uk/damaro/2013/06/07/damaro-project-update-mayjune-2013/

UK Research Data Discovery Service, jisc.ac.uk/rd/projects/uk-research-data-discovery

Page 16
Lewis, S (2013) Thinking about a Data Vault. Edinburgh Research Data blog, datablog.is.ed.ac.uk/2013/12/20/thinking-about-a-data-vault/

Jisc Data Archiving Framework, ja.net/products-services/janet-cloud-services/data-archiving-framework

4. Conclusion

Page 18
Jones, S. Pryor, G. & Whyte, A. (2013) How to Develop Research Data Management Services - a guide for HEIs, DCC How-to Guides, Digital Curation Centre. dcc.ac.uk/resources/how-guides

Jisc Research data management blog. Research at Risk- research data management, where are we now? researchdata.jiscinvolve.org/wp/2014/10/26/research-at-risk-research-data-management-where-are-we-now/

Research Data Management readiness. Meeting the EPSRC policy framework. Workshop February 2015, dcc.ac.uk/events/other-dcc-events/RDM-readiness

Jisc Research Data Spring. Technical tools, software and service solutions to support the researchers’ workflows, the use and management of their data. jisc.ac.uk/rd/projects/research-data-spring

Bibliography

Whyte, A. & Wilson, A (2010) How to Appraise and Select Research Data for Curation, DCC How-to Guides, Digital Curation Centre. dcc.ac.uk/resources/how-guides
Share our vision to make the UK the most digitally advanced education and research nation in the world

jisc.ac.uk