Information

Services

The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Research Data Management Policy Requirements

Support for creating, archiving and sharing research data

Authors: Thomas Parsons, Mark Berry

Date: 04/09/12

Abstract

A summary of Research Data Management requirements from: The University of Nottingham and UK funding council data management policies.

1. Introduction

Research data management has gained prominence in recent years. The Royal Society 'Science as an Open Enterprise' report[1] (June 2012), suggested it should be mandatory to publish research data alongside the traditional research outputs of journal papers and reports. This in turn, supports the Research Councils UK (RCUK) mandate that research data must be made open and available after the research has completed. Alongside these external drivers, Institutional aims and drivers also exist to improve the internal and external processes of sharing data.

Funding from JISC and The University of Nottingham (UoN), will seek to enable these aspects and create an environment for researchers to comply with both funding and Institutional RDM policies. The ADMIRe project was tasked with creating recommendations for a Nottingham preservation and research data collection policy.

Therefore, in order to create a set of technical requirements and make recommendations for future policies, it is important to analyse the drivers for the project and the stakeholders first.

2. Drivers and stakeholders

There are 3 main stakeholders within the ADMIRe project. These are:

- The researchers
- The Institution (UoN)
- The RCUK and other funding bodies

Of these three, the researchers' requirements are being captured through a series of pilots within the ADMIRe project; the Institutions requirements are widespread and are captured through on-going work with departments across the University and are held within policies; while the RCUK requirements are publicly available and widely known.

After initial discussions with researchers and analysis of the UoN RDM survey, it has become clear that few are aware of RDM specific requirements. Hence we will use the pilot studies to capture and challenge requirements derived from the Institution and RCUK mandates. Building on these to ensure we deliver a service that needs all needs.

The first step in creating recommendations for future Nottingham RDM policies is to examine current RCUK and Nottingham policy.

[1] http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

3. RDM Policy Requirements

3.1. Funding requirements

Many of the RCUK funding bodies are mandating that data is actively managed throughout a research project and is then shared at the end. The policies of the key funding bodies are summarised on the DCC website[2], of these EPSRC and Wellcome are typical in their approach:

3.1.1. EPSRC

"In line with the RCUK Common Principles, it is expected that data will be made available in a timely and responsible manner. The EPSRC expects data to be maintained securely for 10+ years."

"Research organisations are expected to publish metadata on the research data they hold, including details of restricted data, outlining the reasons for this and conditions of access. Digital research data should be assigned robust digital object identifiers."

"The EPSRC will monitor progress and compliance on a case by case basis. If it appears that proper sharing of research data is being obstructed, it reserves the right to impose appropriate sanctions."

"The EPSRC does not run a data centre. Research organisations are expected to securely preserve data."

3.1.2. Wellcome Trust

"Institutions are expected to have guidelines setting out responsibilities and procedures for the appropriate storage and disposal of data and samples. Data should be maintained securely for a minimum of 10 years."

"Researchers are to maximise the availability of data with as few restrictions as possible. Data should be made available on publication of results."

"Compliance with the open access policy is monitored.

The Trust seeks to maintain active on-going dialogue with grant holders and provide on-going advice and support for data management and sharing. All awardees are asked to report back on their approach for disseminating their research as part of their end of grant report."

"Data Centre

Where there is no provision, responsibility falls to the institutions in which Wellcome Trust funded researchers are based."

[2] http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

3.1.3. BBSRC

"Data should be made available in a timely and responsible manner. Timely release would otherwise be considered as no later than the release of main findings through publication, or three years as a general guide. Specific scientific areas have established best practice for release of data."

"Data should be kept securely for ten years after the end of a project through research institutions."

"Adherence to the proposed data management and sharing plan will be monitored through the final report assessment procedure and may be taken into account when assessing future proposals."

3.2. Requirements for RDM

The RCUK requirements fall broadly under these areas:

- Data must be released to support a research publication or within a specified time period after the project has finished (BBSRC indicate up to 3 years[3])
- Data must be kept securely for a pre-defined length of time, typically 10 years+
- Data must be actively managed throughout the research project
- The final data set must be deposited within a subject data repository or within an Institutional repository if no subject repository exists
- Data must be available on request
- Research projects must be able to demonstrate compliance to data sharing, this can be on submission of a final project report or throughout the project life

Compliance to these requirements appears to be on a case by case basis. However, there is an implication[4] that projects who can demonstrate compliance, both pre-research and post-research, will be viewed favourably, while those who cannot, risk sanctions. In the case of the BBSRC (a major UoN funder), non-compliance may threaten future funding streams.

Throughout these policies, both the researcher and the Institution are named, thereby placing responsibility for adherence on both parties.

3.3. Requirements from the UoN RDM Policy

Like the RCUK requirements, the UoN Research Data Management policy places an emphasis on both the Principal Investigator (PI) and the Institution to fulfil the policy. There are a number of service and training related obligations that form the basis of ADMIRe WP5 and WP6, and statements that imply a technical infrastructure which falls under WP7. Clause 1.2 and 1.3 states:

[3] http://www.bbsrc.ac.uk/web/FILES/Guidelines/data-sharing-faqs.pdf
[4] http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Data-management-and-sharing/WTX035045.htm#_18._How_will

"1.2. Responsibility for research data management, through a sound research data management plan during any research project or programme lies primarily with Principal Investigators."

"1.3 All new research proposals must include research data management plans or protocols that explicitly address data capture, management, information security (specifically integrity and confidentiality), retention, preservation, sharing and publication."

Here the requirement for a data management plan (DMP) is the responsibility of the PI, yet there is also a management, tracking and a peer-review element to ensure that this requirement is met. How this is handled and tracked is unclear.

Of these, clause 1.5 and 1.6 are explicit in the obligations of the University:

"1.5 The University will provide mechanisms and services for storage, backup, registration, deposit, retention and preservation of research data assets in support of current and future access, during and after completion of research projects."

"1.6 Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University."

These clauses address the requirements mandated by Research Councils UK (RCUK) to:

- Provide the means to store research data during a project
- Preserve research datasets that have been generated using RCUK funding
- Make these datasets available post research project to other researchers
- Create a catalogue or index of research datasets

This in turn, can be broken down into two areas; namely working data and archived data.

3.4. Working research data

The UoN policy suggests that a researcher at the institution will have access to:

- An area they can store their working data and access it
- The mechanisms to backup this data
- The means to make this data available to others if required
- The means to register or catalogue their data

The policy implies that these services are provided by the Institution (i.e. IS) and are not the responsibility of the researcher themselves, although it is not clear whether the Schools are responsible. This is an important point that requires clarification, before a long-term strategy can be formed and a researcher can state how this will be managed within their data management plan.

3.5. Long term archival

Long-term data archival and preservation is a key requirement. Both the UoN policy and the RCUK policies state that research data must be kept securely and be preserved. In the case of the EPSRC, this requirement mandates that a data set will be available to access 10 years from the last time it was accessed. There is no stipulation that the data set should be available online, but the mechanisms should be in place to ensure that this data set will be available in 10 years and it can be requested. This has a number of implications:

- A dataset should be securely archived or preserved, whether digital or non-digital
- The details of every dataset will be available as metadata to aid retrieval and search
- This metadata will be indexed and also preserved long-term
- A method or service will be available to request the dataset
- A mechanism or service exists to allow a researcher to receive the dataset

In the case of EPSRC, 10 years from last access may mean perpetuity. Together with the UoN RDM policy, this implies active management of a dataset and a process to ensure that data which is archived is also preserved long-term. Once a dataset is catalogued and these metadata details are made available, there must be management to ensure the dataset exists in 10 years. Clause 1.5 of the UoN RDM policy firmly places this responsibility for management of datasets with the Institution, while clause 1.6 places this responsibility with the PI if a dataset is stored elsewhere in a subject data repository.

Clause 1.6 acknowledges that subject specific data repositories exist and many are open to researchers with relevant datasets. There is an extensive list of such repositories available from DataCite and this would appear to negate the need to provide an institutional repository or archive for datasets. However, it is widely acknowledged that not all disciplines are served by a subject data repository and finding the correct repository can be time consuming. There are multiple lists of data repositories and not all are open to taking research data of all kinds (2), furthermore, MIT note:

"Not all repositories necessarily take researcher-produced datasets where you can share your data. Moreover, not all repositories listed can ensure long-term preservation of your data; contact each one for more details."

The latter part of this statement is critical to compliance. If the repository is funded through public grants then, as is the case of the Arts and Humanities Data Service[5] or the NBII in the USA[6], it may simply disappear overnight. Therefore, compliance to the Research Councils mandates is bought into question. In this case, a researcher's defence, and principally an Institution's defence, is that the dataset was made available, but due to reasons beyond their control is no longer available. How well this defence will sit with a Research Council is unclear. If Clause 1.6 of the policy has been properly adhered to, a record of the dataset will still exist, thereby drawing attention to this failing.

This scenario also applies to Clause 1.5, if an Institutional data repository or area is created, then precautions should be taken at an Institutional level to ensure the data will be available long-term. This implies access management, digital preservation and a service to control these aspects at an Institutional level.

3.6. Requirements from the Code of Research Conduct and Research Ethics policy

Research data plays a large role in the University of Nottingham Code of Research Conduct and Research Ethics policy document[7]. Clause 4.2 deals specifically with research data:

"4.2 Research data

4.2.1 Data must be recorded in a durable form with appropriate references;

4.2.2 Data must be retained intact for a period of at least seven years from the date of any publication which is based upon them. Data should be stored in their original form – i.e. tapes/discs etc should not be deleted and reused, but kept securely as outlined.

4.2.3 Schools must have procedures for the retention of data. These procedures must be made known to all of their staff and students, who must comply with them.

4.2.4 Confidentiality provisions relating to publications may apply in circumstances where the University or the researcher has made or given confidentiality undertakings to third parties or confidentiality is required to protect intellectual property rights. It is the obligation of the research leader to inform researchers as to whether confidentiality provisions apply and of researchers to enquire of their research leader whether there are any obligations with respect to these provisions."

[5] http://www.ahds.ac.uk/
[6] http://www.nbii.gov/termination/index.html
[7] https://workspace.nottingham.ac.uk/download/attachments/110627444/Code+of+Research+Conduct+and+Research+Ethics+Approved+January+2010.pdf?version=1&modificationDate=1328712590000

Here the emphasis is on both the PI and the Schools to retain data securely, the Institution is not specifically mentioned except in matters of data confidentiality. There is no explanation of how this is to be achieved and Clause 4.2.3 places that responsibility with the School. This implies that each school has their own processes, storage facilities and data management experts to support researchers. Evidence from the survey suggests that 92% of researchers have not received data management training from their School. Therefore it can be assumed that this support is not in place to the level required by RCUK and the University.

Clause 4.2.2 also mandates that research data must be kept securely for 7 years from the date of publication. Responsibility for this appears to lie with the researchers themselves. Obviously, if a researcher leaves the University then there should be a process to manage this at a School level, although anecdotal evidence from the pilots and UoN RDM survey suggests this does not occur.

These requirements are similar to those already identified within the UoN policy and RCUK mandates. The key differences is responsibility, here the responsibility falls on the School or individual researcher rather than the Institution.

3.7. Requirements from the RKTB EPSRC Roadmap

The following statements are derived from "The University of Nottingham Reply to EPSRC Policy Framework on Research Data". They consist of a requirement from EPSRC and the response to this from RKTB.

EPSRC requirements:

"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. "

Response via RKTB:

"These requirements will be addressed by the Private Cloud for Research Data and ADMIRe projects. These projects will develop the infrastructure to host the University's publicly-funded research data holdings and will necessarily require the creation of supporting policies and processes to manage the holdings effectively."

EPSRC requirements:

"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 as available through the DataCite organisation - http://datacite.org)."

Response via RKTB:

"A key output of the ADMIRe project will be to provide mechanisms to deliver this Metadata. Existing widely recognised national and international standards for research metadata and citation will be supported wherever possible."

EPSRC requirements:

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."

Response via RKTB:

"The Private Cloud for Research Data will provide appropriate retention and access controls to meet these requirements."

Again there is an emphasis on UoN to provide the infrastructure to comply with these requirements. ADMIRe will provide the support service infrastructure and metadata capture, while the Private Cloud for Research Data will provide data storage and access control.

The Private Cloud for Research Data was due to be introduced in July 2012 and was expected to utilise FileTek software. Unfortunately this software did not meet expectations and has not been commissioned; therefore the areas that are mentioned in this response require further work.

4. Summary

The policies examined are similar in requirements; the following requirements are mandatory for any over-arching preservation or research data collection policy:

Working data

- Researchers must be provided with a secure area for their working data with access control
- Working research data must be backed up

Deposit

- Research data must be deposited within an Institutional or subject repository in its original digital form. If no subject repository is available, then the funded Institution must provide a data repository to share this data publically.
- Metadata describing each dataset must be created at the time of deposit
- A unique identifier or reference is required to identify the dataset

Retention and preservation

- Data must be preserved in the original digital format for a pre-defined length of time, typically 10 years+
- Data must be accessible throughout a pre-defined time (there is no stipulation as to how it is accessible)
- Metadata must be kept for a pre-defined length of time, typically 10 years+

Administration

- Institutions must be able to demonstrate compliance when asked
- Metadata records should be accessible or restricted to: the researcher, a research group (if applicable), their School, the Institution, the relevant funding bodies or the public