




Research Data Management for librarians

Dublin Core field	
Title	Research Data Management for librarians
Creator	Jones, Sarah
Creator	Guy, Marieke
Creator	Pickton, Miggie
Subject	Research data management; data sharing; research support; librarians; skills.
Description	A course booklet to accompany a three-hour training course for librarians on research data management and how to support researchers.
Publisher	Digital Curation Centre
Date	2013
Type	Text; image
Format	Paper
Identifier	N/A
Source	Content expands on the DCC slideset: 'Research Data Management for librarians'
Language	En
Relation	Format IsBasedOn RoaDMap training booklet <i>Preserving your research data for future use</i> , University of Leeds http://library.leeds.ac.uk/roadmap-project-outputs
Coverage	N/A
Rights	CC-BY  - unless indicated otherwise

Contents

1.	Acknowledgements	2
2.	Tutors.....	2
3.	Course structure	2
4.	Aims and objectives.....	3
5.	Getting started	3
6.	What are research data?	4
	Definitions	4
	Data types	4
7.	Research Data Management	5
	RDM activities	5
	RDM principles and advice.....	6
8.	Data Management Planning.....	8
	Exercise: My DMP – a satire.....	8
	Comparing example DMP responses	10
	Supporting researchers with DMPs	11
9.	Data sharing.....	12
	Benefits and drivers	12
	Exercise: constraints on data sharing	13
	Ethical restrictions.....	16
	IPR and Copyright.....	16
	How to share data.....	16
10.	Skills	17
	How libraries are engaging in RDM.....	17
	Exercise: RDM skills matrix	17
	Potential library RDM roles.....	17
11.	RDM at Northampton.....	19
	Why is research data management important? Drivers for RDM at Northampton.....	19
	History of RDM at the University of Northampton.....	19
	University of Northampton Research Data Management policy.....	19
	University of Northampton Research Data Roadmap	21
	DCC Engagement.....	22
	TUNDRA2 for research data	23
	Exercise: Supporting RDM at Northampton	24
12.	Useful links	25
13.	Annex A - Revised DMP Checklist – draft as of 11/04/2013	26
14.	Annex B - University of Northampton Research Data Policy	30

1. Acknowledgements

This training course was developed as part of the DCC institutional engagement with the University of Northampton. The course has been developed in collaboration with Miggie Pickton, Research Support Librarian at the University of Northampton.

Existing DCC materials were reused, along with content from a number of RDM training courses:

- Barriers to data sharing, by the UK Data Archive, <http://www.data-archive.ac.uk/create-manage/training-resources>
- DIY Research Data MANTRA Training Kit for Librarians, by EDINA and Data Library, University of Edinburgh, <http://datalib.edina.ac.uk/mantra/libtraining.html>
- Managing Your Research Data, by the Research360 project, University of Bath <http://opus.bath.ac.uk/32296>
- RDMRose Lite, by the RDMRose project, University of Sheffield, http://rdmrose.group.shef.ac.uk/?page_id=364
- RoaDMaP training materials, by the RoaDMaP project, University of Leeds <http://library.leeds.ac.uk/roadmap-project-outputs>
- Skills matrix, by the ADMIRE project, University of Nottingham <http://admire.jiscinvolve.org/wp/2012/09/18/rdmnottingham-training-event>
- SupportDM, by the Training for Data Management (TraD) at UEL project, University of East London, <http://www.uel.ac.uk/trad/outputs/resources>

2. Tutors

Sarah Jones, Senior Institutional Support Officer, Digital Curation Centre

Marieke Guy, Institutional Support Officer, Digital Curation Centre

Miggie Pickton, Research Support Librarian, University of Northampton

3. Course structure

The programme for the course is as follows.

- | | |
|----------------------------|---------|
| • Introductions | 15 mins |
| • Research data and RDM | 30 mins |
| • Data Management Planning | 30 mins |
| • Data sharing | 20 mins |
| • BREAK | 15 mins |
| • Skills | 30 mins |
| • RDM at Northampton | 30 mins |
| • Conclusion | 10 mins |

4. Aims and objectives

The aim of the course is to raise awareness of research data management and build confidence amongst librarians to support researchers in this area.

By the end of the session participants will be better able to:

- Understand the varied nature, types and formats of research data
- Describe the activities involved in managing and sharing research data
- Advise on expectations from research funders and the University of Northampton
- Assist researchers with data management and DMPs by pointing to available support

5. Getting started

Recommended pre-course reading is the UKDA guide *Managing and sharing data: best practice for researchers*, available at: <http://data-archive.ac.uk/media/2894/managingsharing.pdf>

If you didn't get chance to look at this in advance, it's a useful resource to consult in the future.

To get started, consider what you understand about research and what you know about managing and sharing data. Do you see a role for librarians in supporting this?

Write down what you want to find out today?

The first question we'll consider is: what are research data? In pairs, list as many types of data as you can, focusing (if appropriate) on the subject areas you support.

6. What are research data?

Definitions

MANTRA provide an explanation of research data in their online training module, *Research Data Explained*, <http://datalib.edina.ac.uk/mantra/researchdataexplained.html>

- Research data are collected, observed or created, for the purposes of analysis to produce and validate original research results.
- Both analogue and digital materials are 'data'
- Digital data can be:
 - created in a digital form ("born digital")
 - converted to a digital form (digitised)

Here are some other definitions of research data:

Queensland University of Technology, Management of Research Data Policy

Research data means data in the form of facts, observations, images, computer program results, recordings, measurements or experiences on which an argument, theory, test or hypothesis, or another research output is based. Data may be numerical, descriptive, visual or tactile. It may be raw, cleaned or processed, and may be held in any format or media. <http://bit.ly/PA21ex>

Cairo Project - Managing Creative Arts Research Data: Training Unit 1

The data might be part of an actual work created through research activity (for example a three-dimensional model displayed via public exhibition) or data may instead be documentary evidence (such as video documentation of a real-world performance event or digital photograph of an installation) of research efforts. Research data includes preparatory, unfinished and supportive work in digital form in addition to data relating to completed works. <http://bit.ly/UHamPG>

EPSRC - Recorded, factual material commonly retained by and accepted in the [research] community **as necessary to validate research findings**; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created. <http://bit.ly/130P5Kv>

Data types

Here is a list of examples of some different types of data that you may encounter:

- Instrument measurements
- Experimental observations
- Still images, video and audio
- Text documents, spreadsheets, databases
- Quantitative data (e.g. household survey data)
- Survey results & interview transcripts
- Simulation data, models & software
- Slides, artefacts, specimens, samples
- Sketches, diaries, lab notebooks ...

7. Research Data Management

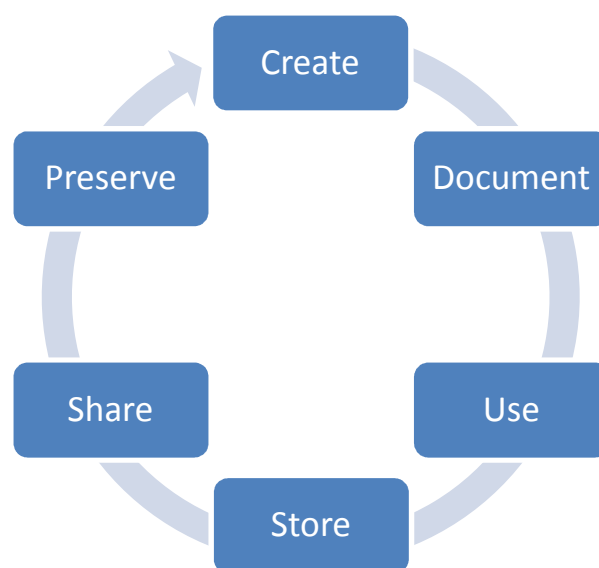
The DCC defines research data management as:

“the active management and appraisal of data over the lifecycle of scholarly and scientific interest”

Research data management is a standard part of good research practice, offering a range of benefits to society, funders of research and the research community. It can help researchers by potentially increasing efficiency, saving time and resources and boosting the impact and visibility of their work.

RDM activities

There are a number of activities involved in managing and sharing data, as seen in the diagram below.



Before creating data, researchers are often encouraged to write data management plans. These plans define how the data will be created, managed, shared and preserved. DMPs are considered in more detail in section 8.

When managing data during the project, researchers should consider how it will be organised so it can be understood and retrieved by the team. This involves considering procedures for filenaming and versioning. When deciding where to store the data, researchers should bear in mind who needs to access it and from where, and how the data will be backed up. Properly managed systems with automated backup, such as central provision from IT services, are more robust and resilient than ‘homegrown’ approaches.

Decisions about data sharing and preservation need to be considered from the outset. If researchers plan to deposit in a data centre, they should check whether there are preferred formats for deposit. Similarly, consent agreements should cover options to share the data at the end of the project if this is planned.

RDM principles and advice

An excellent overview with lots of practical case studies and advice is available in the UKDA guide *Managing and sharing data: best practice for researchers*, available at: <http://data-archive.ac.uk/media/2894/managingsharing.pdf>

Data creation

At this stage researchers should develop procedures to ensure data are created consistently. Filenaming, versioning and structuring of files needs to be considered to make sure the data can be found and understood when needed.

Long-term plans to share or preserve data need to be considered from the outset so that any early decisions don't preclude these options. If researchers plan to share and preserve data:

- Open, standard file formats should be used as they are easier to preserve – some funders may state preferred formats e.g. <http://www.data-archive.ac.uk/create-manage/format/formats-table>
- Consent agreements should also cover options to share and preserve data
- When using third-party data, licences agreements should allow for the derived data to be shared and preserved

Documentation

Contextual information for data is called metadata. This is literally 'data about data'. Metadata and documentation should be created at the time as it is harder to capture later. Certain details can be recorded in a text file (such as a 'read me' file) in the same directory as the data e.g.

- codes for categorical survey responses
- '999 indicates a dummy value in the data'

MIT Libraries provide a useful template for metadata on their Data Management and Publishing website, <http://libraries.mit.edu/guides/subjects/data-management/metadata.html>

Metadata standards should be used where possible to promote reuse and interoperability. The DCC provides a catalogue of disciplinary metadata standards, <http://www.dcc.ac.uk/resources/metadata-standards>

Access and use

Researchers need to be able to access their data whether they are on campus, working from home or are out in the field. They often use multiple devices and should ideally be able to work on a single copy of the data from all of these. They often also need to share data with collaborators who may be based in other universities or outside the Higher Education sector.

Files are often shared via email, transferred on portable storage devices e.g. USB sticks or accessed via Dropbox. There are security risks inherent in this which should be considered to select appropriate methods for filesharing.

Where possible, centrally managed storage and approved tools should be used to provide access to those who need to read or edit the data. The University provides guidance on storage and how to use Netstorage and iFolder to access files remotely, <http://www.northampton.ac.uk/info/20273/it-and-technical/335/storing-your-work>

Storage and backup

Managed services should be used where possible (e.g. University filestores rather than local or external hard drives) in order that data are secure and routinely backed up.

The main principle for backup is:

1. at least 3 copies of a file
2. on at least 2 different media
3. with at least 1 offsite

Data selection

It is not possible for all digital data to be kept forever but outside the archive and library communities there is no widespread recognition of the need to select data for curation. Instead there is a view that “storage is cheap so why don’t we just decide to keep everything”.

The DCC guide *How to appraise and select research data for Curation* offers seven criteria why content may be selected to preserve and share:

1. **Relevance to mission:** The resource content fits the centre’s remit and any priorities stated in the research institution or funding body’s current strategy, including any legal requirement to retain the data beyond its immediate use.
2. **Scientific or historical value:** Is the data scientifically, socially, or culturally significant? Assessing this involves inferring anticipated future use, from evidence of current research and educational value.
3. **Uniqueness:** The extent to which the resource is the only or most complete source of the information that can be derived from it, and whether it is at risk of loss if not accepted, or may be preserved elsewhere.
4. **Potential for redistribution:** The reliability, integrity, and usability of the data files may be determined; these are received in formats that meet designated technical criteria; and Intellectual Property or human subjects issues are addressed.
5. **Non-replicability:** It would not be feasible to replicate the data/resource or doing so would not be financially viable.
6. **Economic case:** Costs may be estimated for managing and preserving the resource, and are justifiable when assessed against evidence of potential future benefits; funding has been secured where appropriate.
7. **Full documentation:** the information necessary to facilitate future discovery, access, and reuse is comprehensive and correct; including metadata on the resource’s provenance and the context of its creation and use.

Preservation

Research funders and institutions have requirements to preserve research data of ‘long-term value’. Typically they expect data to be preserved for 10 years or more. Where possible, researchers should deposit data in subject repositories, data centres and structured databases. See the Databib list for some examples, <http://databib.org> Remember that data centres may have expectations in terms of data quality and format that should be borne in mind from the outset.

8. Data Management Planning

Data management and sharing plans are written at the start of a research project to define how data will be created, managed, shared and preserved. They are often included in grant applications as several research funders require DMPs, but are useful whenever researchers are creating data.

Making a plan helps researchers to save time and effort. By considering what data will be created and how, they can anticipate and avoid problems such as data loss, duplication and security breaches. Planning also enables researchers to make sound decisions, bearing in mind the wider context and consequences of different options. Writing DMPs also encourages researchers to check whether they have the necessary support in place.

The following research funders ask for data management and sharing plans:

● Full Coverage ● Partial Coverage ○ No Coverage

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Access/sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	●	○	●	○	●	●
BBSRC	●	●	●	●	●	●	●	●	●	●	●
CRUK	●	●	●	●	●	●	●	●	●	○	○
EPSRC	●	●	●	●	●	●	●	●	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	●
MRC	●	●	●	●	●	●	○	●	●	○	●
NERC	●	●	●	●	●	●	●	●	●	●	●
STFC	●	●	●	●	●	●	●	●	●	●	●
Wellcome Trust	●	●	●	●	●	●	●	●	●	●	●

From: www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

Funders typically ask researchers to cover the following five themes:

1. Description of data to be collected / created (i.e. content, type, format, volume...)
2. Standards / methodologies for data collection and management
3. Ethics and Intellectual Property (highlight any restrictions on data sharing e.g. embargoes, confidentiality)
4. Plans for data sharing and access (i.e. how, when, to whom)
5. Strategy for long-term preservation

The DCC is currently revising its DMP Checklist and guidance. A draft is available in [Annex A](#).

Exercise: My DMP – a satire

Read through the satirical DMP overleaf. Highlight examples of bad practice and suggest alternative methods / approaches.

My Data Management Plan - a satire

By [C. Titus Brown](#)

Dear NSF,

I am happy to respond to [your request](#) for a 2-page Data Management Plan.

First of all, let me say how enthusiastic I am that you have embraced this new field of "large scale data analysis". Ever since I started working with large Avida data sets in 1993, then with large meteorological data sets in 1995, and then again with large sequence data sets in 1999, I have seen the need for a systematic plan to manage the data. It is nice to see NSF stepping up to the plate in such a timely manner, and I am happy to comply.

Now, as to my actual data management plan, here is how I plan to deal with research data in the future.

I will store all data on at least one, and possibly up to 50, hard drives in my lab. The directory structure will be custom, not self-explanatory, and in no way documented or described. Students working with the data will be encouraged to make their own copies and modify them as they please, in order to ensure that no one can ever figure out what the actual real raw data is. Backups will rarely, if ever, be done.

When required to make the data available by my program manager, my collaborators, and ultimately by law, I will grudgingly do so by placing the raw data on an FTP site, named with UUIDs like 4e283d36-61c4-11df-9a26-edddf420622d. I will under no circumstances make any attempt to provide analysis source code, documentation for formats, or any metadata with the raw data. When requested (and ONLY when requested), I will provide an Excel spreadsheet linking the names to data sets with published results. This spreadsheet will likely be wrong -- but since no one will be able to analyze the data, that won't matter.

Did I mention the click-through license? "You are provided this data for the sole purpose of reproducing our published results. Any attempt to publish your own analyses of this data will be rejected, if necessary during the anonymous review process, by pointing out all of the data cleanup steps you forgot to do correctly in your analysis. (We don't remember all of them ourselves, but there sure were a lot!) Give up now." We will provide a short note -- in a Word document -- detailing the licensing restrictions, as above.

We will make sure that any CSV files we do eventually produce will have format errors, such as missing or extra commas. They will also be encoded in ISO 8859-16, "by accident".

On the off chance that we do choose to provide the source code, it will be in a file named 'source.tar.gz' that unpacks in to the current directory. There will be no explanation of contents, instructions on how to run it, or any enabling information -- it was hard to write, and it should be hard to run! Old, patched, or otherwise impossible-to-obtain versions of Redhat Linux, Perl 5, and associated CPAN libraries will be required before the code runs, even if it doesn't actually use any of them. No source code documentation will be present, of course -- we don't need it ourselves, after all! Automated tests will also not be present (we don't have any of those, either). New versions of the code will be published under the identical file name, with no indication of what changes were made. (We'll be sure to use mixed DOS and Unix EOL editors for our files, so 'diff' won't work to figure out what has changed.)

Note, we didn't use a version control system, either. Or if we did, we made sure to use svn branching and merging profligately, with extremely obscure commit messages (our main programmer only speaks Chinese, so that's how she enters her commit notes. Wouldn't have it any other way). And our repository is not publicly available - you have to beg for permission. Note, I only answer e-mail on every other Tuesday.

Any design notes on the data analysis are in our private e-mail, and we will fight to the death -- up to and including ignoring FOIA requests -- to prevent you from obtaining them.

Meanwhile we will continue publishing exciting sounding (but irreproducible) analyses, and submitting grants based on them, because that's the only thing that the reviewers care about.

sincerely yours,

--titus (representing every computational scientist in the world.)

Comparing example DMP responses

Researchers often ask for examples to help them understand the level of detail required in plans. DMPs should contain enough information for reviewers to be reassured that the plans are robust. However they do not need to go into excessive detail - there is often a strict page limit. The key thing is to ensure the plans are feasible and appropriate. The table below compares the satirical responses from Dr Brown with model answers from ICPSR, NIH and the University of Bristol.

Theme	Satirical response	Example answer
Data organisation	The directory structure will be custom, not self-explanatory, and in no way documented or described.	Data will be stored in a CVS system and checked in and out for purposes of versioning. Variables will use a standardized naming convention consisting of a prefix, root, suffix system. Separate files will be managed for the two kinds of records produced: one file for respondents and another file for children with merging routines specified. [1]
Metadata and documentation	No source code documentation will be present, of course - we don't need it ourselves, after all!	Metadata will be tagged in XML using the Data Documentation Initiative (DDI) format. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively. [1]
Quality assurance	We will make sure that any CSV files we do eventually produce will have format errors, such as missing or extra commas. They will also be encoded in ISO 8859-16, "by accident".	For quantitative data files, the [repository] ensures that missing data codes are defined, that actual data values fall within the range of expected values and that the data are free from wild codes. Processed data files are reviewed by a supervisory staff member before release. [1]
Storage and backup	I will store all data on at least one, and possibly up to 50, hard drives in my lab. Backups will rarely, if ever, be done.	Bristol's Research Data Storage Facility (RDSF) will be used to store the data during the project. The facility represents 2 million pounds of digital resilient storage, with ongoing capital investment. The RDSF is overseen by a steering group of senior research and support staff, which includes the PVC Research. Backup procedures are robust (overnight backup, copies held remotely on tape) and secured access is in place. Recordings made in the field will be copied to the RDSF via a secure web connection, by the PI, as soon as possible, but some delays are expected due to Sri Lankan facilities. Therefore an external, portable hard drive (and an identical copy, for redundancy purpose) will be used for backup in the field. These will remain the responsibility of the PI and filmmaker (each will be carried separately). Checking the quality of recordings made and photographs taken each day will be the responsibility of the PI. [3]

Data sharing	When required to make the data available by my program manager, my collaborators, and ultimately by law, I will grudgingly do so by placing the raw data on an FTP site, named with UUIDs like 4e283d36-61c4-11df-9a26-edddf420622d. I will under no circumstances make any attempt to provide analysis source code, documentation for formats, or any metadata with the raw data. When requested (and ONLY when requested), I will provide an Excel spreadsheet linking the names to data sets with published results. This spreadsheet will likely be wrong - but since no one will be able to analyze the data, that won't matter.	Data products from this study will be made available without cost to researchers and analysts. User registration is required in order to access or download files. As part of the registration process, users must agree to the conditions of use governing access to the public release data, including restrictions against attempting to identify study participants, destruction of the data after analyses are completed, reporting responsibilities, restrictions on redistribution of the data to third parties, and proper acknowledgement of the data resource. Registered users will receive user support, as well as information related to errors in the data, future releases, workshops, and publication lists. [2]
Restrictions on sharing	Any design notes on the data analysis are in our private e-mail, and we will fight to the death - up to and including ignoring FOI requests - to prevent you from obtaining them.	The data will be deposited with [repository] but not disseminated for one year to give the investigators time to publish their findings. [1]

Sources used for example answers:

- [1] ICPSR Framework for a Data Management Plan <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/framework.html>
- [2] NIH Examples of data sharing plans http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#ex
- [3] University of Bristol Example AHRC Technical Plan <http://data.bris.ac.uk/files/2013/02/data.bris-AHRC-Technical-Plan-v21.pdf>

Supporting researchers with DMPs

There are lots of things that you may offer to support researchers with DMPs. These include templates on what to include in plans, guidance and examples, or even libraries of successful DMPs for researchers to browse and reuse. In support cases researchers may require more hands-on support via training courses or consultancy services. The University of Northampton has also customised the DMPonline tool and directs researchers to use this.

Some useful advice to share with researchers is to:

- Keep DMPs short and simple – lots of detail is not required
- Be specific – name standards and quantify the scale of the data
- Seek advice – consult and collaborate to develop sound plans
- Base plans on available skills and support to ensure implementation is feasible
- Justify any resources requested and explain proposed restrictions on sharing. These are allowed when reasonable.

9. Data sharing

Data sharing is:

“... the practice of making data used for scholarly research available to others.” [Wikipedia]

It involves: the data sharer; the data repository; the secondary data user and support staff!

Benefits and drivers

There are lots of benefits for researchers. Data sharing:

- Helps to avoid duplication, thereby reducing costs and wasted effort
- Promotes scientific integrity and debate
- Enables scrutiny of research findings and allows for validation of results
- Leads to new collaborations between data users and data creators
- Improves research and leads to better science
- Increases citations*

* A study by Piwowar, Day and Fridsma showed a 69% increase in citation,

<http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000308>

There are also a number of drivers for data sharing. The RCUK Common Principles on Data Policy state:

“Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.”

<http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

Many individual research funders have research data management requirements – the Digital Curation Centre has collated a useful summary of Research Council policies at,

<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

The University of Northampton is committed to making the research outputs of its members freely accessible to the wider community. The University maintains an open access digital repository, a collection of online journals and a research data policy which promotes open access to research data.

You can hear more about open access from the Vice Chancellor and other academics in some videos on open research, <http://www.northampton.ac.uk/info/20137/research/806/open-research-at-the-university-of-northampton/2>

Exercise: constraints on data sharing

List some reasons why certain data can't be shared and consider whether any actions could be taken to reduce or overcome these restrictions.

Constraints on data sharing	Possible solutions / approaches

UKDA potential barriers to data sharing – with suggested solutions



Available from <http://data-archive.ac.uk/create-manage/training-resources>

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
1	My data is not of interest or use to anyone else.	It is! Researchers want to access data from all kinds of studies, methodologies and disciplines. It is very difficult to predict which data may be important for future research. Who would have thought that amateur gardener's diaries would one day provide essential data for climate change research? Your data may also be essential for teaching purposes. Sharing is not just about archiving your data but about sharing them amongst colleagues.
2	I want to publish my work before anyone else sees my data.	Data sharing will not stand in the way of you first using your data for your publications. Most research funders allow you some period of sole use, but also want timely sharing. Also remember that you have already been working with your data for some time so you undoubtedly know the data better than anyone coming to use them afresh. If you are still concerned you can embargo your data for a specific period of time.
3	I have not got the time or money to prepare data for sharing	It is important to plan data management early in the research data lifecycle. Data management ideally becomes an integral part of your research practice, reduces time and financial costs and greatly enhancing the quality of the data for your use too.
4	If I ask my respondents for consent to share their data then they will not agree to participate in the study.	Don't assume that participants will not participate because data sharing is discussed. Talk to them – they may be less reluctant than you might think, or less concerned over data sharing! Make it clear that it is entirely their decision, whereby they can decide whether their data can be shared, independent of them participating in the research. Explain clearly what data sharing means, and why it may be important. But they are still free to consent or not. You can always explain what data archiving means in practice for their data. If you have not asked permission to share data during the research, then you can always return to gain retrospective permission from participants.
5	I am doing highly sensitive research. I cannot possibly make my data available for others to see.	The first thing is to ask respondents and see if you can get consent for sharing in the first instance. Anonymisation procedures can help to protect identifying information. If these first two strategies are not appropriate then consider controlling access to the data or embargoing for a period of time. Also data that is held in the UK Data Archive is not publically available. Only registered researchers can gain access to the data.

6	I am doing quantitative research and the combination of my variables discloses my participant's identity.	Quantitative data can be anonymised through processes of aggregation, top coding, removal of variables, or controlled access to certain variables (i.e. postcodes).
7	I have collected audiovisual data and I cannot anonymise them, therefore I cannot share these data.	Visual data can be anonymised through blurring faces or distorting voices, but this can be time consuming and costly to carry out. It can mean losing much of the value of the data. It is better to ask for consent to share data from participants in an unanonymised form, and/or control access to the data.
8	I have made promises to destroy my data once the project finishes.	Why were such promises made? Always avoid making unnecessary promises to destroy data. There is usually no legal or ethical need to do so, except in the case of personal data. But that certainly would not apply to research data in general. Also consider where you have received this advice from? You may need to negotiate with research ethics committee or ethics boards about this agreement.
9	My data have been gathered under complete assurances of confidentiality.	Again why was such an assurance made? It is best to avoid unnecessary promises. Anonymisation procedures can be implemented to protect identities, but confidentiality can never be completely guaranteed. You can also consider controlling access to the data.
10	My data collection and resulting transcripts are in a foreign language.	This should not be a problem. The UK Data Archive can accept foreign language transcripts although translations into English are preferred.
11	It is impossible to anonymise my transcripts as too much useful information is lost.	Get in touch with us at the UK Data Archive. We may be able to help and it might not be as difficult as it looks. Also, access controls on the data may be a better solution than anonymisation if too much useful information would be lost.
12	My data collection contains data which I have purchased and it cannot be made public.	It is important to know who holds the copyright to the data you are using and to obtain the relevant permissions. You need to be aware of the licence conditions of the data you are using and what you can and cannot do with the data.
13	Other researchers would not understand my data at all - or may use them for the wrong purpose.	Producing good documentation and providing contextual information for your research project should enable other researchers to correctly use and understand your data.
14	There is IPR in the data.	This should not be a problem if you seek copyright permission from the owner of the intellectual property rights. This is best done early on in the research project, but could be sought retrospectively.

Ethical restrictions

There are many benefits to publishing research data. However there is also legislation such as the Data Protection Act and commercial confidentiality that demands care is taken when publishing data. Some research may be commercially valuable and covered by confidentiality agreements. Research data that compromises such areas of confidentiality should not be published.

The UKDA suggests that it is often possible to make data available if three key issues are addressed:

- Informed consent, addressing both current and future use.
- Confidentiality – data may need to be anonymised so that individuals, organisations or businesses cannot be identified.
- Access control – it may be necessary to control who has access to data. This may enable data sharing for research and educational purposes.

Further information is available at, <http://data-archive.ac.uk/create-manage/consent-ethics>

IPR and Copyright

Ownership of the data should be carefully considered, particularly if the research utilises data from third party sources who own copyright. Ideally, ownership should be clarified at the beginning of the research process. If the rights in the data are unclear, this may limit what can be done.

For data where you own the rights – or where you have appropriate permissions from third-party copyright holders – you should consider what licence or end user agreement will be applied if your data is made available to others. For example, will it be appropriate to apply a global licence such as a Creative Commons or Open Data Common Licence? Guidance on how to licence research data is available from the DCC: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>

How to share data

Data are often shared informally between researchers or may, for example, be made available from a project web site. Locating an appropriate data repository can be an excellent way to share data and to make sure that it will be available in the longer term. Repositories are likely to apply some quality control standards to the data and may, for example, assign a unique identifier to the data, making it more readily findable and citable. Always check the terms of conditions of any repository service. A list of data repositories is available from Databib at, <http://databib.org>

To give data the best chance of sustainability, it is better to use formats that are open, documented standards rather than proprietary or patented formats. Higher quality, uncompressed formats are better to use for master files. Some recommended formats are listed below and online at, <http://www.data-archive.ac.uk/create-manage/format/formats-table>

Type	Recommended	Avoid for data sharing
Tabular data	CSV, TSV, SPSS portable	Excel
Text	Plain text, HTML, RTF, PDF/A only if layout matters	Word
Media	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	Quicktime, H264
Images	TIFF, JPEG2000, PNG	GIF, JPG
Structured data	XML, RDF	RDBMS

10. Skills

How libraries are engaging in RDM

The DCC is running a series of Institutional Engagements to support universities to develop RDM services, <http://www.dcc.ac.uk/community/institutional-engagements>

Based on the 21 universities that we are working with, we have found that the library leads RDM initiatives in the vast majority of cases. Research offices, IT teams and other services such as records management are also involved, but often to a lesser degree. University libraries have been instrumental in many areas of work, including making the case to senior management, defining the institutional strategy and delivering RDM training and support services.

Librarians are well-placed to support researchers given their highly relevant skillset. Good existing relationships with researchers also make it easier to provide support. The open access leadership roles of libraries often means they are expected to address RDM too.

Exercise: RDM skills matrix

Based on the activities involved in RDM that we discussed earlier, consider who may have relevant skills or expertise to share. Complete the table overleaf by ticking in the box when support for the given topic could realistically be provided by that group.

Potential library RDM roles

There are many roles that library staff may take on to support researchers with RDM, including:

- Leading on local (institutional) data policy
- Bringing data into undergraduate research-based learning
- Teaching data literacy to postgraduate students
- Developing researcher data awareness
- Providing advice, e.g. on writing DMPs or advice on RDM within a project
- Explaining the impact of sharing data, and how to cite data
- Signposting who in the Uni to consult in relation to a particular question
- Auditing to identify data sets for archiving or RDM needs
- Developing and managing access to data collections
- Documenting what datasets an institution holds
- Developing local data curation capacity
- Promoting data reuse by making known what is available

Consider which roles you may feel confident to support. Is this an exciting opportunity or a frightening prospect given already overloaded schedules? Ideas and concerns will be discussed further in the next section when we consider RDM support at Northampton.

Research Data Management activity	Library and Learning Services	IT Services	Other professional services
Authentication/access authorisation			
Copyright			
Current awareness of the research data landscape			
Data citation			
Data information literacy			
Data migration (to maintain access)			
Data security (to control access and to avoid data loss)			
Data storage			
Digital preservation			
IP, patents, commercialisation of research data			
Licensing (of UoN's and other datasets)			
Metadata and documentation			
Repository management			
Research costing			
Research data and publications			
Research Data Management Policy			
Research funders' compliance			
Research integrity and governance (ethics & consent)			

11. RDM at Northampton

Why is research data management important? Drivers for RDM at Northampton

The drivers for RDM at Northampton are exactly the same as those at other universities:

- As part of the REF, universities are being asked to demonstrate that their **research environment** is fit to support high quality research. The existence of services and support for RDM is a key component of a strong research environment.
- Pressure from **funders** is growing all the time. Previously funders responsible for distributing public monies asked only that research publications should be made available open access, now they expect researchers to share their own datasets, and **re-use and re-purpose** those of others. Funders ask for formal data management plans and expect the sector to provide the services and infrastructure to support these.
- In times of increasing accountability, **publishers** in some fields are demanding access to the data which underpin published work. These data must be accessible and meaningful to others so that work can be verified and replicated.
- Also on the theme of accountability, universities are subject now to FOI and EIR requests and of course to the constraints of Data Protection **legislation**. With clear evidence that data are stored and managed responsibly the university will be less likely to suffer **reputational** damage. With appropriate data management planning it may even be possible to delay release of research data until the needs of the researcher are met.
- Last but not least, good RDM is simply **good research practice** – being organised, documenting datasets and curating them to ensure long term access are all important practices that will produce long term benefits to the original researcher and others.

History of RDM at the University of Northampton

The Research Support team in LLS has been working towards addressing researchers data management needs since mid 2010. The poster on the next page (which was presented to a JISC/DCC workshop in October 2012) summarises this work.

The poster is available in NECTAR: <http://nectar.northampton.ac.uk/4573>

University of Northampton Research Data Management policy

The university's research data policy was put together by a small working group comprising a research manager, a researcher, the Records Manager and Research Support Librarian. It was prompted by recommendations from the LLS Research data (DAF) project and sought to draw attention to the issue of formal research data management and raise awareness of good practice. The DAF project report is available at <http://nectar.northampton.ac.uk/2736>

Northampton was one of the first universities to approve a research data policy (see <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-institutional-data-policies>) and it is probably one of the shortest, choosing to confine itself to following the good practice recommended by RCUK (see the RCUK policy and code of conduct on the governance of good research conduct, pp. 7-8). A supporting procedure accompanies the policy.

Research Data Management at the University of Northampton



About the University of Northampton

- University status achieved in 2005
- Increased focus on research and supporting the research community
- 200+ research students plus more research active staff
- Responding to sector wide drivers: REF; data publication; funders...

The challenge

Back in 2010:

- Little was known centrally about university researchers' data storage requirements, or the research workflow that incorporates the creation and management of data
- No university wide data storage policy or procedure existed
- In NECTAR (our institutional repository), we had available the infrastructure to store and preserve digital data
- Previous studies had noted that the process of undertaking a **DAF project** had been valuable in itself, even if the resulting inventory of data was only partial



Data Asset Framework (DAF) project

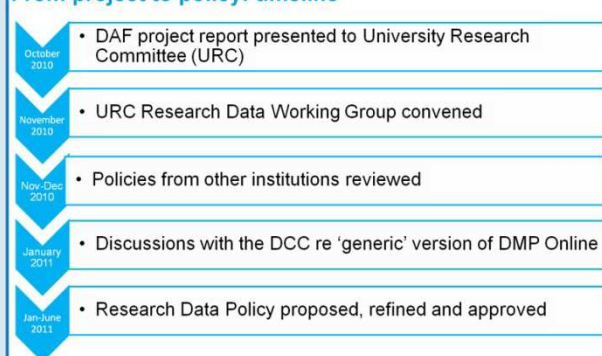
- Project team:
 - Library led, with two project researchers (graduate interns) plus a Project Board comprising staff with expertise in repositories, records management and collection development
- Ran from May to June 2010 (eight weeks)
- Data collection, three stages:
 - Initial interviews with research leaders in each School; online survey of researchers; one-to-one interviews with researchers
- Topics covered:
 - Types, sizes and formats of research data; data ownership; storage; security; sharing and access (short and long term); funders' requirements



DAF project outcomes

- 80 researchers responded to the survey and 16 agreed to take part in the follow-up interviews; all Schools were represented
- Some common behaviours were identified but also variance
- Data storage needs, behaviours and vulnerabilities varied through the research lifecycle.
- 56% of participants agreed that they would like a university repository to store their research data, but not necessarily to offer open access
- Nine recommendations were made, covering:
 - Reporting to senior research managers and leaders
 - Creation of a university research data policy and procedure
 - Clarification of the ownership of research data
 - Training and guidance
 - Dissemination of findings

From project to policy: timeline



The university's research data policy and procedure¹

The University of Northampton expects researchers to:

- Follow the RCUK (Research Councils UK) code of good research conduct² with respect to research data:
 - Keep clear and accurate records of research procedures and results
 - Hold records securely
 - Ensure data are accessible to others after the completion of the research (depositing in a national collection if possible)
 - Manage data in accordance with funders' policies and relevant legislation
- Produce and follow a data management plan (DMPOnline³ is recommended)

In exchange the university will:

- Provide a dedicated storage facility for research data
- Offer training and support for research data management

We have also worked with researchers and members of the university's Research and Strategic Bidding Office to develop a **research data roadmap** in response to the EPSRC's call for action.

¹ <https://www.northampton.ac.uk/info/20283/academic-research/1606/research-data-policy>

² <http://www.rcuk.ac.uk/documents/reviews/grc/goodresearchconductcode.pdf>, pages 7-8

³ <https://dmponline.dcc.ac.uk/>



Working with the DCC

As part of our institutional engagement we are working with Sarah Jones and Marieke Guy of the DCC to:

- Implement the university's research data policy:
 - Raise awareness of good data management practice
 - Offer training and advice in writing data management plans
 - Create a university storage facility for research data
- Establish researchers' data management needs and develop appropriate training and services to support these, including:
 - One to one research data 'clinics'
 - Subject specific training for Schools and research groups
 - Targeted training for research students and library staff
- Promote the university's research data roadmap at the highest level to win support for its recommendations

The policy was presented twice to University Research Committee, the first time receiving several objections:

- Duplication of effort - “we have to do this already ” ... for funders, professional bodies, etc.
- Relevance or applicability to different disciplines
- Reluctance to set disposal date (or even review date) – “I’d be very upset if my data were deleted” ...after I had left
- Aversion to procedures being mandatory
- Expense – who will pay for it?

The policy that was eventually approved took a softer approach, *encouraging* researchers to complete data management plans rather than *mandating* them to do so. It no longer required DMPs for all research projects and it suggested simplified internal procedures. A default five year review period satisfied both researchers and the Records Manager.

The policy is replicated in Annex B.

University of Northampton Research Data Roadmap

In March 2011 the EPSRC announced its new policy framework on research data:

<http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx> The policy framework followed the RCUK Common Principles on Data Policy <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>.

The EPSRC required their funded institutions to comply with the policy framework and to demonstrate this compliance by developing a roadmap “*to align their policies and processes with EPSRC’s expectations by 1st May 2012 and to be fully compliant with these expectations by 1st May 2015*” (<http://www.epsrc.ac.uk/about/standards/researchdata/Pages/impact.aspx>).

Although Northampton does not currently receive funding from EPSRC, it has won funds in the past, and aspires to do so in future. Moreover, it seems likely that other research councils will follow the EPSRC’s lead. It was therefore felt that a Northampton roadmap should be developed. Another working group was formed, this time comprising representatives from Science and Technology, the RSBO and LLS.

The roadmap (short extract shown overleaf) describes how the University will meet the EPSRC’s requirements concerning the management of research data. It has been written with the EPSRC requirements in mind but represents good practice for all research areas. It maps current and planned practice to the EPSRC expectations and covers awareness of the regulatory environment; connection with published papers; access to datasets; use of metadata; and data curation.

The roadmap was presented to University R&EC in April 2012 and approved in principle. However, since full implementation of the roadmap requires additional resource, further approval was required from the University’s Executive Team. Since this is still outstanding the roadmap has not been released.

Expectation	What do we do now and what do we have in place?	What must we do to meet any identified gaps?	When will we do it?	What resources will we commit?
1. 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.	<p>New research students receive basic training in research data management (RDM) as part of their mandatory induction to the university.</p> <p>Records management, Freedom of Information and Data Protection are covered in staff development sessions which are open to all staff and research students.</p>	<p>Provide tailored RDM awareness and training opportunities to researchers of all disciplines.</p> <p>Data Protection training to be mandatory for all researchers dealing with personal data.</p>	Summer 2013	Staff from LLS and Records Management will work with experts from the Digital Curation Centre (DCC) to develop and deliver appropriate further training sessions and materials. These will subsequently be made available as part of the existing Staff Development Programme.

DCC Engagement

Since mid 2012 we have been working with the DCC on one of their 21 institutional engagements, <http://www.dcc.ac.uk/community/institutional-engagements>. These provide collaborating institutions with the support they need to increase research data management capability.

Having already made significant progress by establishing researchers' data needs, defining a policy and creating a roadmap, we chose to focus efforts on its implementation. In particular, we asked the DCC to help with training and guidance for both researchers and those who support them.

So far Sarah and Marieke have run **training** sessions on:

- Managing your PhD data (for research students)
- Managing data through the research lifecycle (Northampton Business School)
- Meeting funders' requirements for RDM (Social Sciences)

And provided **guidance**:

- Creation of a DMPonline template for the University of Northampton, with attached guidelines
- Development of a guide to meeting ESRC data management planning requirements (in conjunction with John Horton)

We have also run one-to-one RDM clinics for researchers

Still to come:

- Further training for Schools (please encourage them to ask for this)
- A series of posts on the Research Support Hub, <http://researchsupporthub.northampton.ac.uk>
- Further support for research data storage...

TUNDRA2 for research data

Under the leadership of Phil Oakman, the University is rolling out TUNDRA2 as an institution-wide content management system for storing, managing and preserving files. The expectation is that TUNDRA2 will eventually replace networked storage drives (such as the R: drive) as the place where staff will store their files.

TUNDRA2 offers greatly enhanced functionality for file management and collaborative work, and for this reason is a suitable tool for managing research data.

Jane Callaghan, Judith Sixsmith and colleagues are currently piloting the service for their major EU-funded research project . TUNDRA2 allows them to store and remotely access research data, financial information, meetings minutes, published outputs and so forth all in the same place. Access controls are available right down to the level of the individual file so documents and data can be shared among only those with appropriate permissions. The software also incorporates communication tools such as instant messaging and blogging – a great advantage for an international project involving researchers from four countries.

We hope to use this pilot to increase our understanding of researchers' actual requirements and, ideally, to inform the development of a project template in TUNDRA2 which can be used by all sizes of research project.

If you are aware of any projects coming up that might find TUNDRA2 useful then do let Phil or Miggie know.

Exercise: Supporting RDM at Northampton

- In small groups, discuss which RDM activities you think should fall within your role and which shouldn't.
- Do you feel confident to support RDM?
- How would you like to see things develop?
- You have 15 minutes

If needed, please consult the 'potential library RDM roles' section on p17 for ideas

Activities that fall within your role	What help do you need to enable you to feel confident in supporting this activity?

12. Useful links

RDM policies and requirements

Overview of funders' data policies, www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

Summary of funder requirements for DMPs, <http://www.dcc.ac.uk/resources/data-management-plans>

RCUK Common Principles on Data Policy, www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

University of Northampton Research Data Policy, (Also given in full in Annex B)
<http://www.northampton.ac.uk/info/20283/academic-research/1606/research-data-policy>

Guidance, tools and resources

Databib list of research data repositories, <http://databib.org>

DCC How-to guides, www.dcc.ac.uk/resources/how-guides (series includes how to develop a DMP, select data, licence data, cite data and develop RDM support services)

DMPonline, a tool for writing DMPs <https://dmponline.dcc.ac.uk>

ICPSR Framework for Creating a Data Management Plan,
<http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/framework.html>

Massachusetts Institute of Technology (MIT) Libraries, Data Management and Publishing website,
<http://libraries.mit.edu/guides/subjects/data-management/index.html>

UK Data Archive, Managing and sharing data: best practice for researchers, <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

Training materials

DIY Training Kit for Librarians, <http://datalib.edina.ac.uk/mantra/libtraining.html>

RDMRose Lite for librarians, http://rdmrose.group.shef.ac.uk/?page_id=364

Research Data MANTRA, <http://datalib.edina.ac.uk/mantra>

Research360 RDM presentation for PhD students, <http://opus.bath.ac.uk/32296>

RoadMaP training, <http://library.leeds.ac.uk/roadmap-project-outputs>

SupportDM online modules for librarians, <http://www.uel.ac.uk/trad/outputs/resources>

UKDA training resources, <http://www.data-archive.ac.uk/create-manage/training-resources>

13. Annex A - Revised DMP Checklist – draft as of 11/04/2013

ADMINISTRATIVE INFORMATION

ID	<i>A pertinent ID as determined by the funder and/or institution</i>
Researcher / PI	<i>Name of main researcher(s) or Principal Investigator(s) on the project</i>
Project data contact	<i>Name (if different to above), telephone and email contact details</i>
Project name	<i>If applying for funding, state the name exactly as in the grant proposal</i>
Project description	<i>Brief summary to explain the purposes for which the data are being collected</i>
Funder	<i>State research funder if relevant</i>
Date	<i>Date of creation or update</i>

DATA COLLECTION

1. What data will you collect or create?

- What type, format and volume of data?
- Do the chosen formats and software enable sharing and long-term validity of data?
- Are there any existing data that you can reuse?

Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data volumes in terms of storage, backup and access.

2. How will the data be created?

- What standards or methodologies will you use?
- How will you structure and name your folders and files?
- How will you handle versioning?
- What quality assurance processes will you adopt?

Outline how the data will be collected/generated and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. And explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.

DOCUMENTATION AND METADATA

3. What documentation and metadata will accompany the data?

- What information is needed for the data to be to be read and interpreted in the future?
- How will you capture / create this documentation and metadata?
- What metadata standards will you use and why?
- How will potential users find out about your data?

Describe the types of documentation that will accompany the data to provide a secondary users with any necessary details to prevent misues, misinterpretation or confusion. This may include details on the methodology used, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data. Consider how you will capture this information and where it will be recorded. You are strongly encouraged to utilise community standards where they exist.

ETHICS AND LEGAL COMPLIANCE

4. How will you manage any ethical issues?

- How will you protect the identity of participants if required? e.g. via anonymisation
- Have you gained consent for data preservation and sharing?
- How will sensitive data be handled to ensure it is stored and transferred securely?

Ethical issues affect how you store data, who can see/use it and how long it is kept. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you're aware of this and have planned accordingly. If you're carrying out research involving human participants you must also ensure that consent is obtained to share data.

5. How will you manage copyright and intellectual property ownership issues? (IPR)

- Who owns the data?
- How will the data be licensed for reuse?
- Are there any restrictions on the reuse of third-party data?
- Will data sharing be postponed / restricted e.g. to seek patents?

State who will own the copyright and IPR of any new data that you will generate. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Also consider permissions to reuse third-party data and any restrictions needed on data sharing.

STORAGE AND BACKUP

6. How will the data be stored and backed up?

- Do you have sufficient storage or will you need to include charges for additional services?
- How will the data be backed up?
- Who will be responsible for back-up and recovery?

Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage provided by university IT teams is preferable. Similarly, it is normally better to use automatic backup services provided by IT Services than rely on manual processes.

7. How will you manage access and security?

- How will you control access to keep the data secure?
- What are the risks to data security and how will these be managed?

If your data is sensitive (e.g. detailed personal data, politically sensitive information or trade secrets), you should discuss any appropriate security measures and note any formal standards that you will comply with e.g. ISO 27001.

SELECTION AND PRESERVATION

8. Which data are of long-term value and should be shared and/or preserved?

- How will you decide what to keep?
- How long will the data be retained and preserved?
- What are the foreseeable research uses for the data?

Consider how the data may be re-used e.g. to validate your research findings, in new studies from pooled datasets or in teaching. Decide which data to keep and for how long. This could be based on any obligations to retain certain data, the potential reuse value, what is economically viable to keep, and any additional effort required to prepare the data for data sharing and preservation.

9. What is the long-term preservation plan for the dataset?

- Where (e.g. in which repository) will the data be held?
- Will charges be applied by data repositories?
- Have you costed in time and effort to prepare the data for sharing / preservation?

Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.

DATA SHARING

10. How will you share the data?

- How will users be able to find and access the data? e.g. via a repository
- With whom will you share the data, and under what conditions?
- When will you make the data available?

Consider where, how, and to whom the data should be made available. Will you share data via a data repository, handle data requests directly or use another mechanism? The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Mention earlier examples to show a track record of effective data sharing.

11. Are any restrictions on data sharing required?

- What action will you take to overcome or minimise restrictions?
- For how long do you need exclusive use of the data and why?
- Will a data sharing agreement (or equivalent) be required?

Outline any expected difficulties in data sharing, along with causes and possible measures to overcome these. Restrictions to data sharing may be due to participant confidentiality, consent agreements or IPR.

RESPONSIBILITIES AND RESOURCES

12. Who will be responsible for data management?

- Who is responsible for each data management activity?
- How are responsibilities split across partner sites in collaborative research projects?

Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Individuals should be named where possible.

13. What resources will you require to deliver your plan?

- Is additional specialist expertise (or training for existing staff) required?
- Do you require hardware or software which is additional or exceptional to conventional desk-based research and institutional provision?
- Will charges be applied by data repositories?

Carefully consider any resources needed to deliver the plan. Where dedicated resources are needed, these should be outlined and justified.

14. Annex B - University of Northampton Research Data Policy

Available at: <http://www.northampton.ac.uk/info/20283/academic-research/1606/research-data-policy>

In June 2011 the University's Research Committee approved the following policy and procedure for the management of research data.

The University of Northampton recognises that good research is underpinned by good research data management.

In accordance with the recommendations of Research Councils UK, the university expects researchers to:

- "Keep clear and accurate records of the research procedures followed and the results obtained, including interim results
- Hold records securely in paper or electronic form
- Make relevant primary data and research evidence accessible to others for reasonable periods after the completion of the research: data should normally be preserved and accessible for [at least] ten years"...
- ..."Manage data according to the research funder's data policy, and all relevant legislation
- Wherever possible, deposit data permanently within a national collection." (Research Councils UK, 2009, p.5)

If no appropriate national collection exists then following the completion of the research project all data will be deposited in a secure central storage facility to be provided by the university.

In order to meet these expectations, the Principal Investigator is, at an early stage of their research project, encouraged to produce and then follow a data management plan (DMP). The [DMP Online](#) tool is recommended for this purpose.

Research data management procedure

At the start of every funded project, the Principal Investigator will produce a DMP in line with the funder's specific requirements, or if no requirements are specified by the funder, then in accordance with the good practice described by [DMP Online](#).

The Principal Investigator will be responsible for ensuring that the actions outlined in the DMP are carried out.

At the end of a project data will be deposited in either a national collection or in the university's secure central storage facility. If the latter then the data must incorporate adequate descriptive metadata and, if necessary, accompanying explanatory documentation.

Datasets within the university's central storage facility will be subject to review after five years following deposit. The review will be undertaken by the researcher and the Records Manager. In the absence of the original researcher, an appropriate academic colleague will conduct the review with the Records Manager.

If required as a result of the review, data will be destroyed or otherwise securely disposed of by the University Records Manager.

Support for research data management

The following support is currently available to researchers:

- Help with identifying national data archives for the external storage of research data
- Professional advice and support for researchers who receive Freedom of Information (FOI) or Environmental Information Regulation (EIR) requests. Note that if you are a member of University staff and you receive such a request you should notify the University Records Manager immediately.
- Training and advice on research data management

A central storage facility for research data is coming soon.

For further information about research data services at The University of Northampton, please call Miggie Pickton, Research Support Librarian or Phil Oakman, University Records Manager.