

Data Management Planning

STFC FUNDING APPLICANTS



Science & Technology
Facilities Council

Author	Andrew Burnham
Date	02 September 2015 (original 15 July 2013)
Version	1-3
URL	http://www.le.ac.uk/researchdata
File	UoL_DMP_STFCGuide_v1-3
Reviewed by	Research Infrastructure Advisory Group, Research Data Management Working Group, IT Services, Information Assurance Services.
IPR	
Classification	May be viewed by anyone, anywhere in the world
Acknowledgement	<i>This document is an adaptation of a document produced by the data.bris team at the University of Bristol, whose permission to adapt the document is gratefully acknowledged.</i>

Document History

Version	Date	Author	Detail/Reason for Change
1-3	02.09.2015	A. Burnham	Inclusion of RDM Principles, updates and corrections
1-2	27.03.2014	A. Burnham	Minor edit.
1-1	20.02.2014	A. Burnham	Minor edit.
First published version (1-0) \UoL_DMP_STFCGuide_v1-0.docx (.pdf)	15.07.2013	A. Burnham	First published version.
Draft 4 \UoL_DMP_STFCGuide_v0-4.docx	08.07.2013	A. Burnham	Final draft with minor edits.
Draft 3 \UoL_DMP_STFCGuide_v0-3.docx	06.06.2013	A. Burnham	Third draft for review from RCMG. Approved for release by PVC Research
Draft 2 \UoL_DMP_STFCGuide_v0-2.docx	30.05.2013	A. Burnham	Second draft for review by RCMG.
Draft 1 \UoL_DMP_STFCGuide_v0-1.docx	22.04.2013	A. Burnham/University of Bristol	First draft using University of Bristol document as basis.



The Scientific and Technology Facilities Council (STFC) adopts the view that 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.

1. Summary of STFC data requirements

The following summarises STFC data requirements:

- STFC-funded research data must be made freely available after project completion.
- A defined period of exclusive data use is permitted.
- A data management plan (DMP) is typically required at the proposal stage, explaining how the STFC requirement will be met.
- Applicants should ensure that raw data remains available for *ten years* from project completion, while data which is not re-measurable is retained '*in perpetuity*'.
- Data underpinning published research outputs should be available within six months of the output's publication.

2. Introduction to data requirements

For the STFC, research data is a major output of funding and a major source of economic impact. Scientific publishers, too, increasingly require that data which underpins a published research output (a journal article for instance) should be made available for validation purposes. STFC data requirements apply to all scientific research data produced using STFC funding, including grants paid to universities, access to beam time at STFC-supported facilities and (through STFC subscriptions) to other organisations such as CERN or ESO.

The DMP is usually an integral part of any grant application made to the STFC.

3. Where to get help and information

Refer to the University research data website www.le.ac.uk/researchdata where specific funder related information and the latest data management advice will be included.

The range of appropriate contacts includes:

- IT Services
- Library
- Research Support Office
- Leicester Learning Institute
- Information Assurance Services
- Enterprise and Business Development

A single point of contact is also available: email researchdata@le.ac.uk at any time and as early as possible in the bid process. This will means specific queries or general request for assistance can be directed to the right place(s). You can also request assistance with development of a data management plan via this email address.

It is also recommended that you use the Digital Curation Centre (DCC) DMPOnline¹ resource to create a data management plan (DMP) using the STFC template and requirements. As and when University of Leicester templates and specific guidance are created this will be confirmed on the RDM website².

Specific research IT services available include Research File Storage, high performance computing, Wiki, 'LAMP' stack (a general purpose, Linux, relational database and web hosting service, based around open source software- Linux, Apache, MySQL and PHP), file transfer (FileDrop) and source code control (Subversion SVN)³.

In 2014 the University agreed it's **RDM Principles**⁴ which act to guide researchers and inform funders of the University approach (see 10. below).

STFC and general Information

STFC Scientific Data Policy (2011)	http://www.stfc.ac.uk/info/scientific-data-policy/
STFC Data Management Plan	http://www.stfc.ac.uk/funding/research-grants/data-management-plan/
Digital Curation Centre STFC Funder's Data resource	http://www.dcc.ac.uk/resources/policy-and-legal/research-funding-policies/stfc
Digital Curation Centre 'DMP Online' tool	https://dmponline.dcc.ac.uk/
Digital Curation Centre DMP Checklist	http://www.dcc.ac.uk/resources/data-management-plans/checklist
UK Data Archive – Managing and Sharing Data	http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
University Research Data Management Website	www.le.ac.uk/researchdata
University Research Data Management Support	researchdata@le.ac.uk
University Research Data Management Principles	http://www2.le.ac.uk/services/research-data/documents/uol_rdmprinciples
RCUK Joint Electronic Submission System (Je-S)	https://je-s.rcuk.ac.uk/JeS2WebLoginSite/Login.aspx

¹ DMPOnline, <https://dmponline.dcc.ac.uk/>

² Data Management Planning, <http://www2.le.ac.uk/services/research-data/create-data/DMPlan>

³ IT Services, <http://www2.le.ac.uk/offices/ithelp/>

⁴ RDM Principles, http://www2.le.ac.uk/services/research-data/documents/uol_rdmprinciples

4. What is research data?

STFC's definition of research data covers:

- a) *Raw* scientific data directly arising as a result of experiment/measurement/observation.
- b) *Derived* data which has been subject to some form of standard or automated data reduction procedure.
- c) *Published* data, i.e. data which underpins a publication and from which scientific conclusions have been derived.

Where possible STFC expects raw data to be retained for a minimum period, which is often years after project completion. Where raw data cannot be retained (due to size considerations, for example) the processes used to create derived data should be documented in detail. This documentation should then be available alongside derived data.

Efforts should be made to retain any research data which cannot be re-measured 'in perpetuity'.

Software as a form of data in its own right is excluded from STFC requirements. Research outputs (such as journal articles) are also excluded. There may also be exceptional reasons why data cannot be made publicly available (for example, legislation or ethical issues). If you believe this applies to your research, this should be explained in your Data Management Plan (DMP) at the funding application stage.

5. The Data Management Plan

The DMP is usually an integral part of any grant application made to the STFC: "*Proposals for projects that would result in the production or collection of scientific data should include, in the case for support, a data management plan.*"⁵

STFC does not stipulate a specific format for data management plans but recommends that eight areas be addressed, asks that it is "*clear, concise and proportionate*" and takes into account the "*value of the data, the likely level of future re-use, the complexity and levels of data, and the cost effectiveness of preservation compared to re-creation*"⁶:

1. Specify the types of data the research will generate.
2. Specify which data will be preserved and how.
3. Specify the software and metadata implications.
4. Specify for how long the data will be preserved.
5. Specify and justify which data will have value to others and should be shared.
6. Specify and justify the length of any proprietary period.
7. Specify how data will be shared.
8. Specify and justify any resources required to preserve and share the data.

⁵ STFC, Data Management Plan, <http://www.stfc.ac.uk/funding/research-grants/data-management-plan/>

⁶ STFC, Data Management Plan

STFC advises that Digital Curation Centre (DCC)⁷ guidance is followed. The DCC provides an on-line DMP tool (DMP Online)⁸ which provides the ability to create a DMP according to the specific requirements of various funding councils.

The DMP should be submitted alongside your main Je-S application. Every STFC-supported facility is also expected to have an on-going and regularly updated DMP of its own. Your DMP should explain how you'll manage any research data that you plan to generate or collect, in line with STFC's requirements. An assessment of the DMP will be made as part of the general assessment of your application. Your DMP should follow best practice in information management.

6. Data formats

As part of your DMP you should state in which format/s your data will be collected, analysed and stored (for example, Open Document Format, CSV file or Excel spreadsheet). When selecting a data format to use, your own research needs must come first. If you find you do need to use a non-standard format, you should consider converting your data to a more widely usable format once your own data analysis is complete. If you're unsure which file formats to use, the UK Data Archive publishes a list of recommended formats for deposit⁹. These formats may also be appropriate for non-specialised uses during your research.

7. Describing data

Metadata¹⁰ is 'data about data' and is information (or cataloguing information) that enables data users to find and/or use a dataset. In your DMP you should outline plans for documenting your research data, to meet both your own needs and those of later users. Descriptions of your data could be kept in a separate, dedicated database.

You should also outline in your DMP how you'll name files and folders to make sure you and others can access the data easily, for example by keeping track of different versions.

In cases where no discipline-specific metadata standards exist, it may help to imagine another data user attempting to make sense of your data in your absence, after your project has concluded. If presented with only the data itself, such a user may be faced with the difficult task of 'unpicking' it. How will they make sense of your file and folder naming conventions? What extra information would they need to make maximum use of your data? How will they understand how derived data was created from raw data?

8. Providing access to data

The STFC require all research data to be made freely available after project completion, although a period of exclusive use by the data's originators is acceptable, to allow results to be exploited. The length, nature and reason for a period of exclusive access should be given in your DMP. *Published*

⁷ Digital Curation Centre, <http://www.dcc.ac.uk/>

⁸ DMP Online, <https://dmponline.dcc.ac.uk/>

⁹ UK Data Archive File Formats Table, www.data-archive.ac.uk/create-manage/format/formats-table

¹⁰ Metadata, <http://www2.le.ac.uk/services/research-data/organise-data/metadata>

data (data underpinning a published research output) should be available within six months of the output's publication, in order to allow scientific findings to be independently verified.

It is expected that data sharing will be achieved via an appropriate repository (or even several different repositories if a number of different datasets are generated). Such repositories should be named by the applicant within their DMP.

The University of Leicester is currently (2015) in the process of developing a research data repository.

Repositories should be selected in order to 'maximise the scientific value' obtained from the aggregation of research data. In areas where such national or international repositories exist, it is strongly recommended that applicants include in the DMP references to the repository's archiving policies. The DMP should list the planned datasets (and accompanying metadata), provide a brief description of each one and say for how long the repository will make them available, and under what terms. Making data 'freely available' may be achieved by means of a user registration process, i.e. data needn't be openly available to anonymous users.

The University requires that all research publications are deposited with the Leicester Research Archive (LRA)¹¹.

If no appropriate repository exists, an applicant's DMP is likely to be much more detailed and should describe plans for offering sustained access to research data, including the minimum availability period of ten years (for raw data).

9. Citing research data in research outputs

All journal articles and conference proceedings submitted for publication after 1 April 2013 which are the result of RCUK funded research must be made available for anyone to read without charge (made available on 'open access'). Open access means that anyone with an internet connection can read your research paper or conference proceeding without the frustration of hitting a subscription or publisher paywall. The benefits in terms of wider dissemination, greater openness and transparency, and speeding up of discovery are considerable.

This requirement includes providing a means by which third parties can access any underpinning research datasets. This may be a reference (such as a unique URL or DOI) printed in a paper, which will lead an enquirer to a specific web page where the data is available. The enquirer might be directed to a page which displays the contact details of a custodian of the data, whom they are asked to email in order to gain access to the data.

Given the extended timescales involved in this process (possibly extending beyond the mandatory three years mentioned above), it is strongly recommended that the authors of published academic outputs *do not provide their current contact details* as a means of accessing underpinning research data, as these details will change over time. If you plan to use an established data repository service, ask this service for a unique reference identifier which could be included in the publication

¹¹ Leicester Research Archive, <http://www2.le.ac.uk/library/for/researchers/publish/open-access>

instead. If you're not planning to use an established data repository service, contact researchdata@le.ac.uk or the Library for further guidance.

10. University RDM Principles

In 2014 the University agreed its **RDM Principles**¹² which act to guide researchers and inform funders of the University approach and should be referred to in funding proposals.

Research data are defined as any material created or collected for the purposes of analysis to generate and validate original research results, irrespective of the format of data. Research data may be digital, paper based or in other forms. Examples of different types of research data include datasets, images, text (such as transcripts of interviews), audio and video recordings, and computer scripts.

Scope

1. *These principles apply to all research conducted at the University, regardless of funding source. They do not imply additional compliance where good practice and relevant research funders' requirements are already being followed.*

Research inception and planning

2. *Data management planning is an integral, essential and dynamic component of the research process from inception and should include provision for the selective long term custodianship of research data.*
3. *Research proposals should include all possible recovery of direct costs of research data management where the funder allows this.*

During the research: management and storage of data

4. *During the research process, data are an asset which needs to be appropriately managed and stored: to meet legislative, funder, information governance and University requirements; to facilitate data security (confidentiality, integrity, availability); to facilitate appropriate access, collaboration and sharing of data and results.*
5. *Data can be actively managed throughout, following and updating the data plan, recognising that storage and its funding is not infinite, with ongoing decisions regarding retention and destruction.*

¹² RDM Principles, http://www2.le.ac.uk/services/research-data/documents/uol_rdmprinciples

After the research: retention, sharing, publishing, citation, re-use

6. *When the research has been completed, research data (including websites) of long term value, or data required by funders or the University must be selected for retention, then preserved and curated for as long as appropriate.*
7. *Data retained in these circumstances must be offered to funder or discipline repositories and/or to the UK Web Archive as appropriate. If such repositories are unavailable or unsuitable, data must be stored in a University repository. Data deposited with external repositories or unsuitable for making open access must be registered with the University.*
8. *There is a presumption of open access to data held in a University or other public repository. However, access may be restricted, subject to a time embargo or not permitted for legal (i.e. intellectual property, data protection, confidentiality, contractual requirements), ethical or commercial reasons.*
9. *Data should not be deposited with any organisation that does not commit to appropriate access and availability for re-use and exclusive rights to re-use or publish research should not be handed to commercial publishers, unless this is a condition of funding.*
10. *The re-use or sharing of data that are made available should not be unnecessarily restricted by licences or terms of use.*
11. *All research outputs must cite data produced and/or used during research as appropriate, detailing access to that data.*

Responsibilities

12. *Primary accountability for research data management lies with the most senior University researcher associated with the work or project. Responsibility for research data management may be delegated.*
13. *During the research process, researchers are responsible for adherence to legal requirements such as Data Protection and for the creation of metadata and other documentation that enables data to be discoverable, understandable and re-useable.*
14. *After the deposit of data with a repository, the repository is responsible for the on-going management of that data in accordance with legal, technical and other requirements.*
15. *The University will be responsible for providing a Research Data Management service led by the Library to include training, advice, guidance and data curation.*
16. *The University will secure sustainable solutions that meet the requirements for long term data storage and re-use as set out in these principles.*

The Managing Research Data guide series comprises:

- An Introduction to Managing Research Data – For Researchers and Students
- Data Management Planning – AHRC funding applicants
- Data Management Planning – BBSRC funding applicants
- Data Management Planning – EPSRC funding applicants
- Data Management Planning – ESRC funding applicants
- Data Management Planning – MRC funding applicants
- Data Management Planning – NERC funding applicants
- Data Management Planning – STFC funding applicants
- Data Management Planning – Non-RCUK funding applicants

They are part of a range of RDM material produced by the University, all available via www.le.ac.uk/researchdata.



What would you do if you lost your research data tomorrow?

Take the research data health check... and find help to secure, share and exploit your valuable research.

Chances are you could use some helpful pointers in all of these!

Create	Organise	Keep	Find & Share
 <p>Have you...</p> <ul style="list-style-type: none"> <input type="checkbox"/> fully understood your research funders' data management requirements? Government and bodies require that publicly funded research is made available for reuse – or you go in debt with their grant policies? Your future funding might depend on it! <input type="checkbox"/> written a data management plan? Your funder may already require this to be built in from the original stage to avoid headaches in the future. <input type="checkbox"/> gained ethics approval/consent? Writing a data management plan will aid planning and help you to manage ethics and governance requirements. <input type="checkbox"/> protected your intellectual property? Saving intellectual property considerations for a rainy day could be the appropriate next, alongside other projects and perhaps your through future health! 	 <p>Are your research files and data...</p> <ul style="list-style-type: none"> <input type="checkbox"/> clearly described, in terms of content being standard metadata? Are you still heavily 'in there' in terms of data being lost? Will you be able to remember how you generated your data, and will you or anyone else be able to find it in the future when you wish to reuse and share? <input type="checkbox"/> clearly labelled with versions and dates? Have you remembered which was the relevant version and which dataset was used in producing a peer research outcome? <input type="checkbox"/> legally structured and named? Can you be confident that you generated data, can you still find the dataset too? <input type="checkbox"/> future-proofed against broken links, losing persistent identifiers? The persistent identification of digital resources can play a vital role in enabling their accessibility and usability over time using recommended data standards. 	 <p>Do you know...</p> <ul style="list-style-type: none"> <input type="checkbox"/> how to restrict access to your research data to the right people? Have you considered with authority or data centre experts that only the right people have access to your research? <input type="checkbox"/> which data to keep and which data to discard? Managing research data effectively means being selective, which data to discard and when as well as what to keep and to share long! <input type="checkbox"/> how securely your data is stored? What happens if your storage media fail? How robust is it? Can't get it all in the first and most sensible place? <input type="checkbox"/> how your data is backed-up? Have you made use of university and/or external resources to back up data so that you have multiple copies in case of loss or theft? 	 <p>Do you know how to...</p> <ul style="list-style-type: none"> <input type="checkbox"/> find existing information resources related to your research? Where can you find research data that you can re-purpose or combine with your own to produce new research? <input type="checkbox"/> share data with your collaborators securely and effectively? Whether building a collaborative project, generating results for others to comment on or sharing the final results of your research – how do you link with your colleagues ahead of water sharing? <input type="checkbox"/> deposit your research data and outputs to an open repository? Is there an appropriate disciplinary or institutional repository and should you need to be to deposit your research output? How should the deposit be managed? <input type="checkbox"/> publish your research, and get it cited as well? Institutions and data centres must make research data available to others while ensuring credit to the researcher who did the work. Your future career could depend on it!

To find information, support, advice and training, as well as links to external resources, go to www.le.ac.uk/researchdata or email: researchdata@le.ac.uk