University of Leicester

Data Management Planning

BBSRC FUNDING APPLICANTS

BBSRC bioscience for the future

www.le.ac.uk/researchdata

Author	Andrew Burnham
Date	02 September 2015 (original 15 July 2013)
Version	1-3
URL	http://www.le.ac.uk/researchdata
File	UoL_DMP_BBSRCGuide_v1-3.pdf
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	*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.*

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_BBSRCGuide_v1-0.docx (.pdf)	15.07.2013	A. Burnham	First published version.
Draft 4 \UoL_DMP_BBSRCGuide_v0-4.docx	08.07.2013	A. Burnham	Final draft with minor edits.
Draft 3 \UoL_DMP_BBSRCGuide_v0-3.docx	06.06.2013	A. Burnham	Third draft from review by RCMG. Approved for release by PVC Research.
Draft 2 \UoL_DMP_BBSRCGuide_v0-2.docx	28.05.2013	A. Burnham	Second draft for review by RCMG.
Draft 1 \UoL_DMP_BBSRCGuide_v0-1.docx	23.04.2013	A. Burnham/University of Bristol	First draft using University of Bristol document as basis.

The Biotechnology and Biological Sciences Research Council (BBSRC) supports the view that publicly-funded research data is a public good, produced in the public interest and should be openly available to the maximum extent possible. The re-use of data can lead to new scientific understanding.

Holders of BBSRC grants are encouraged and expected to follow BBSRC Policy[1] in order to practise and promote data sharing and create a scientific culture within which data sharing is embedded.

This guide is designed for BBSRC applicants who are required to submit a Data Sharing Statement along with their application.

1. Summary of BBSRC data requirements

- Data must be released no later than the publication of findings and within three years of project completion.
- Commercial interests should not "*unduly delay or prevent sharing*" of data funded by BBSRC.
- Data must be available for a minimum of ten years after project end and in a form appropriate for secondary use.
- A Data Management plan is required for all research proposals submitted.
- Sharing via an appropriate, established repository is expected in research areas where such repositories exist.
- Applicants for responsive research funding should complete a 'Data Sharing Plan' as part of their research grant proposal in order to demonstrate a willingness to share data.
- Compliance with the BBSRC Data Sharing Policy is checked by BBSRC as part of the Final Report Assessment at the close of a project.
- Certain data types are considered by the BBSRC to have an especially high re-use value.
- Funding to support the management and sharing of research data can be requested as part of the full economic cost of a research project.

[1] BBSRC Data Sharing Policy, www.bbsrc.ac.uk/datasharing

2. What you need to know - introduction to data requirements

Holders of responsive mode research grants must observe the BBSRC data sharing policy. Applicants for responsive research funding streams should complete a 'Data Sharing Plan' as part of their research grant proposal in order to demonstrate a willingness to share data.

The BBSRC Data Sharing Policy does not currently apply to studentships and fellowships (though it is planned to extend it further). BBSRC-funded students and research fellows are instead 'encouraged to consider the policy' when making applications.

In the case of responsive mode grant holders, compliance is checked by BBSRC as part of the Final Report Assessment at the close of a project.

The BBSRC Data Sharing Policy also applies to institutions in receipt of core strategic grants and researchers funded by such institutions. The monitoring of institutional compliance is carried out as part of the Institute Assessment Exercise.

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) DMPOnline[2] resource to create a data management plan (DMP) using the BBSRC template and requirements. As and when University of Leicester templates and specific guidance are created this will be confirmed on the RDM website[3].

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)[4].

[2] DMPOnline, https://dmponline.dcc.ac.uk/
[3] Data Management Planning, http://www2.le.ac.uk/services/research-data/create-data/DMPlan
[4] IT Services, http://www2.le.ac.uk/offices/ithelp/

In 2014 the University agreed it's **RDM Principles**[5] which act to guide researchers and inform funders of the University approach (see 9. below).

BBSRC and general Information

BBSRC Data Sharing Policy (with FAQs and Data Sharing leaflet)	http://www.bbsrc.ac.uk/about/policies/position/policy/data-sharing-policy/
BBSRC Impact Policy	http://www.bbsrc.ac.uk/about/policies/position/policy/impact-policy/
European Bioinformatics Institute, BioModels Database	http://www.ebi.ac.uk/biomodels-main/
UK Data Archive File Formats Table	http://www.data-archive.ac.uk/create-manage/format/formats-table
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
Digital Curation Centre BBSRC Funder's Data resource	http://www.dcc.ac.uk/resources/policy-and-legal/research-funding-policies/bbsrc
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
RCUK Joint Electronic Submission System (Je-S)	https://je-s.rcuk.ac.uk/JeS2WebLoginSite/Login.aspx

4. Exceptions to data sharing

BBSRC acknowledges that in exceptional circumstances data sharing may not be possible or may not be desirable. If you believe this applies to your research, you should use your Data Sharing Statement as an opportunity to explain why you feel this is the case. You may feel the costs of sharing data with anticipated low re-use would make data sharing not worthwhile.

Note that only in rare circumstances will BBSRC accept that ethical considerations preclude the sharing of *all* the research data you generate. It is far more likely that some additional actions (such as anonymisation of data and seeking appropriate permissions) may be required in order to facilitate re-use of the data.

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

BBSRC recognises the need to protect opportunities for commercialisation of research outputs, but they state that commercial interests should not "*unduly delay or prevent sharing*" of data funded by BBSRC.

5. Research data with a high re-use value

Added value can be gained by using certain datasets for purposes other than those for which they were originally designed. Researchers in all areas are encouraged to consider data sharing where there is scientific merit in doing so. The following types of research data are however, considered by the BBSRC to have a high re-use potential:

Data from high volume experiments:

Datasets containing hundreds of measurements generated in parallel from a single experimental sample (for example, omics, sequencing etc.).

Low throughput data from long time series or cumulative approaches:

Standardised measurements collected at regular intervals forming a resource that can be subjected to retrospective analysis. Often this type of data is of particularly high value as it cannot be substituted or replaced.

Models generated using systems approaches:

Models created using iterated systems approaches can be as important as the data they generate. They are sharable and re-usable assets with a high re-use potential. Such models should be freely available to any researchers wishing to reproduce experiments.

An appropriate repository in which to deposit models is the BioModels Database[6]. Authors are encouraged to submit models to the BioModels Database before publication of an associated paper (depositors will receive an identifier that they can use in the publication), but models will only be publicly available on the BioModels Database once the paper has been published.

6. Costs of sharing research data

BBSRC recognises that data sharing has time and cost implications. Funding to support the management and sharing of research data (for example, staffing and physical resources such as storage and networking capability) can be requested as part of the full economic cost of a research project.

7. The 'Data Sharing Statement'

In order to demonstrate an intention to comply with the BBSRC Data Sharing Policy, applicants are required to complete a Data Sharing Statement at the time of application. An additional page of the Case for Support is allocated for this purpose. Using this page for any other purpose will result in automatic rejection of the proposal.

[6] BioModels Database, http://www.ebi.ac.uk/biomodels-main/

Your Data Sharing Statement will be assessed by peer reviewers separately from the rest of the proposal. However, if an inappropriate Data Sharing Statement is submitted, an applicant's credibility will suffer. Proposals with exceptional scientific value but a poor Data Sharing Statement may be offered a conditional award; alternatively, a redrafted Data Sharing Statement may be requested. In individual cases, BBSRC reserves the right to take a more prescriptive approach to data sharing.

It is suggested that applicants use the following headings in their Data Sharing Statement:

Data types:

An estimate of the volume and type of data that you expect to generate (for example, experimental measurements, models or images). BBSRC encourages researchers to outline *all* research data and *not only that which directly underpins a research publication.*

Data and metadata standards:

In order to maximise re-use value, BBSRC researchers should generate and manage data using widely accepted formats and methodologies. In some disciplines these are well defined; in others, standards are still being developed.

An example of an established standard is the widespread use of models in SBML and CellML formats. The BioModels Database only accepts models in these formats in order to help to verify the reproducibility of results.

Numerous conversion tools exist (such as SBFC), which will transform models in other formats into SBML or CellML models. If you find you do need to use a non-standard technology during the course of your research, consider standardising datasets and models prior to sharing them.

Where no clear disciplinary guidelines exist concerning which formats 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 re-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 deposit formats[7]. These formats may also be appropriate for use throughout your research.

A major barrier to data sharing is the widespread use of non-standard, highly specialised file formats. In order to make use of data, a number of digital technologies must be available, which are known as technological 'dependencies'. These may be fairly common technologies such as a desktop PC, the Windows 7 operating system and Adobe Reader 9 software. Also the technology required to access data might be rare and hard to acquire, or even unique. You should address this problem by minimising the number of technological dependencies involved in using your data as much as possible.

Where dependencies are inevitable you should favour 'open' technologies rather than proprietary ones. Proprietary technologies are owned by a vendor or group of vendors. Commercial pressures

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

may lead to the withdrawal of a particular piece of hardware or software, in favour of a new and possibly incompatible replacement. In contrast, 'open' technologies are supported by a community of users and do not have the same commercial vulnerabilities.

Metadata[8] is 'data about data' or 'cataloguing information' that enables data users to find and/or use a digital output. In your Statement you should briefly outline plans for documentation, both to meet your own needs (i.e. to ensure that you can find what you want, when you need it) and those of later users.

Where no clear disciplinary metadata standards as yet exist, it may help to imagine another user attempting to make sense of your output in your absence. If presented with only the data, they may be faced with the difficult task of 'unpicking' it. So, for example, how would they make sense of file and folder naming conventions? Has any special software been used in the creation of an output that must also be available in order to use it? How was secondary date derived from primary data?

Relationships:

Relationships to any other datasets available for re-use.

Secondary uses:

Briefly describe the re-use potential which the dataset or model will have, once complete.

Data sharing:

BBSRC recognises two broad approaches to data sharing: 1) via an established, online public repository and 2) by the award holder providing data on request. A combination of these two approaches may be appropriate and either of them may be subject to specific access mechanisms (such as a requirement to have a data sharing agreement) in order to protect confidential or otherwise sensitive data.

Sharing via an appropriate, established repository is expected in research areas where such repositories exist.

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

Note that the responsibility for ensuring data is retained remains with the grant holder even when data is deposited with an external archive for sharing purposes. BBSRC therefore recommends that a copy is retained locally for security purposes.

Researchers have the option of providing data directly on request to third parties. Researchers choosing to do so should however, consider the BBSRC requirement to make data available for a minimum of ten years after the project ends, in a format appropriate for secondary use. Some updating of both data formats and accompanying metadata is to be expected during this period.

[8] Metadata, http://www2.le.ac.uk/services/research-data/organise-data/metadata

BBSRC suggests that any data preparation which is required before sharing (such as standardisation or quality checking) should be done within the lifetime of the funded project, in order to avoid subsequent loss of staff or motivation.

Data sharing timeframe:
Some communities have established time frames for releasing data (for example, the Crystallography Protein Data Bank, where a twelve-month delay between publishing the first paper on a structure and making co-ordinates public for secondary use is typical, or Metabolomics [MeT-RO] where a six-month delay in publication can be requested). It is the responsibility of the applicant to reference such disciplinary guidelines in the Data Sharing Statement.

Where no clear community guidelines exist, BBSRC expects data to be released *no later than the publication of findings* and *within three years of project completion.*

Primary data should be securely retained, in an accessible format, for a minimum of ten years after project completion.

Proprietary data:
Specify any data which will not be freely available for re-use and a brief explanation of why this is the case e.g. a co-funder may have requirements which conflict with those of the BBSRC.

Format/s:
List of the formats of final datasets or models (see Data and metadata standards, above).

8. Citing research data in research outputs

From 1st April 2013 all the UK's research funding councils, as part of RCUK, require research outputs i.e. journal articles to provide 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. Alternatively 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 instead.

9. University RDM Principles

In 2014 the University agreed its **RDM Principles**[9] 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.*

[9] 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.