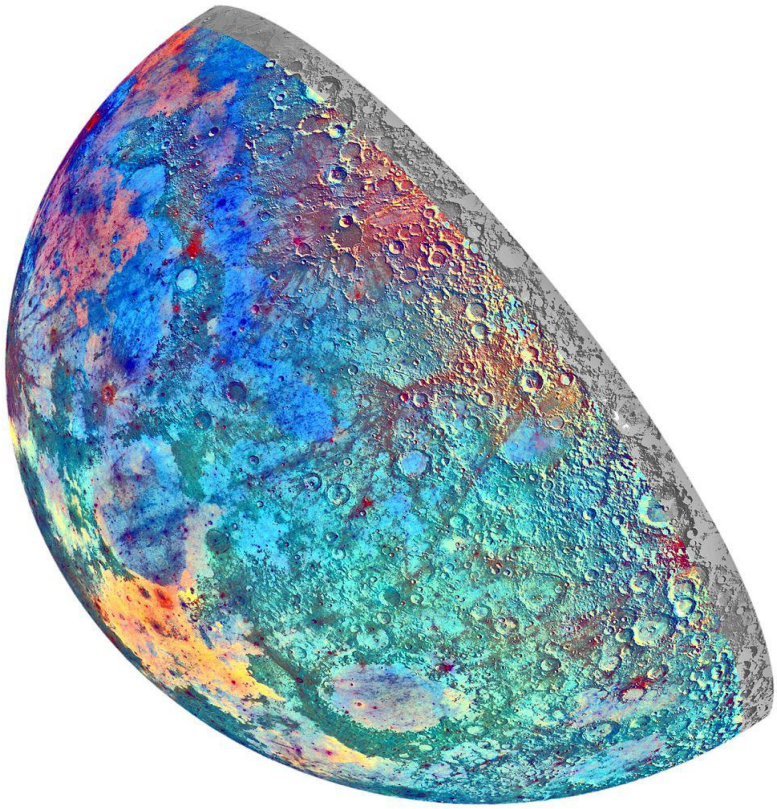
Data Management Planning

STFC funding applicants

Version 1.6 June 2021



University of Bristol

Research Data Service

Image: 1024px-Moon Crescent False Color Mosaic, NASA/JPL, CC-BY-SA 2.0

University of BRISTOL

SUMMARY

- A DMP is typically required at the proposal stage, explaining how the STFC requirement will be met.
- STFC-funded research data must be made freely available after project completion.
- A defined period of exclusive data use is permitted.
- 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.

INTRODUCTION

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.

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.

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 ten 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.

The Data Management Plan

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

alongside your main Je-S application. Every STFC-supported facility is also expected to have an ongoing and regularly updated Data Management Plan 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[1]. 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.

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 deposit.[2] These formats may also be appropriate for non-specialised uses during your research.

Describing data

Metadata 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 a secondary 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 secondary 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?

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 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

[1] STFC guidance on Data Management, https://stfc.ukri.org/funding/research-grants/data-management-plan/

[2] UK Data Archive Recommended file formats table, www.ukdataservice.ac.uk/manage-data/format/recommended-formats

repositories if a number of different datasets are generated). Such repositories should be named by the applicant within their DMP.

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 references to the repository's archiving policies in the DMP. 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 of Bristol has its own research data repository which researchers from any discipline may wish to use. This repository can provide ongoing access to research data for extended periods of time and issue unique DOIs for deposited datasets (see citing research data in research outputs, below). For smaller datasets, no costs are involved. If you are planning to deposit larger datasets with the repository a cost may be incurred. Contact the Research Data Service[3] as early as possible if you believe you'll need to make use of Bristol's data repository.

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).

CITING RESEARCH DATA IN RESEARCH OUTPUTS

From 1st April 2013 all the UK's research funding councils, as part of UKRI (formerly 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 within 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 displaying the contact details of a custodian of the data and asked to email them in order to gain access to the data.

Given the extended timescales involved in this process (possibly extending beyond the mandatory ten years mentioned above) it is strongly recommended that the authors of published academic outputs do not provide their current contact details as a means by which underpinning research data may be accessed as these will change over time. If you plan to make use of an established data repository service, ask it for a unique reference identifier which might be included in the publication instead. If you're not planning to use an established data repository, contact the Research Data Service[3] for further guidance.

[3] http://www.bristol.ac.uk/staff/researchers/data/contacts/