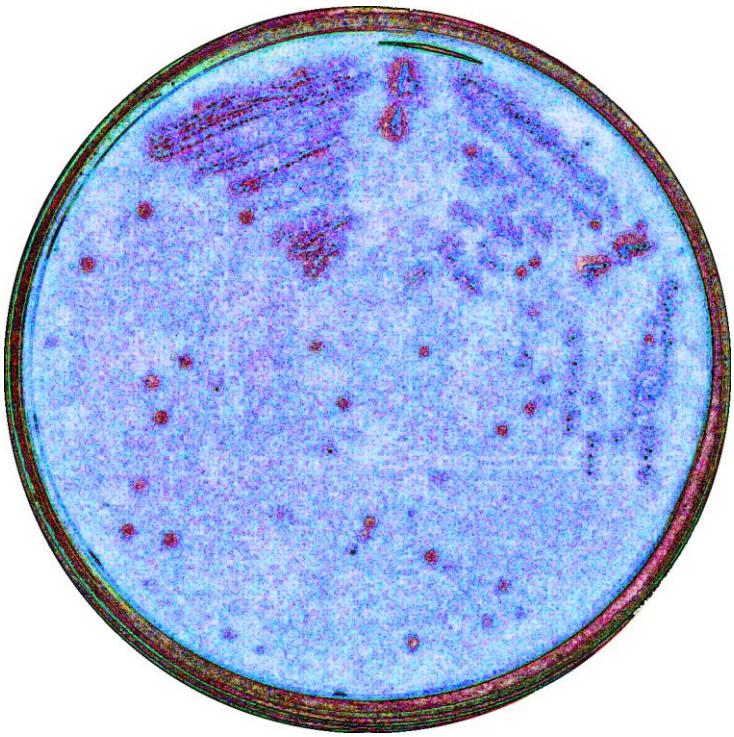
Data Management Planning

# MRC and NC3Rs funding applicants

Version 1.9 June 2021



University of Bristol

## Research Data Service

Image: Streptococcus pyogenes-inoculated trypticase soy agar, CDC/ Richard R. Facklam, Public Domain

University of BRISTOL

# SUMMARY

The MRC expects valuable data arising from MRC-funded research to be made available to the scientific community with as few restrictions as possible, so as to maximize the value of the data for research and for eventual patient and public benefit. Such data must be shared in a timely and responsible manner.

The following points are intended to serve as a reminder of MRC data requirements:

- All MRC-funded researchers have a responsibility to ensure that opportunities for data re-use are maximised, within the regulatory requirements of the law.
- Applicants are typically required to submit a Data Management Plan (DMP) at the application stage. A template is provided by the MRC.[1]
    - A simple DMP could be fewer than 500 words long;
    - In cases where rich resources are to be created, a DMP should be longer, up to a maximum of 3 x A4 pages long;
    - Population- or patient-based studies must meet twenty-one additional requirements;
    - Established population- or patient-based studies should already have a Policy on Data Sharing which can be referred to in your DMP;
    - Applicants proposing new population- or patient-based studies should contact the Research Data Service before submitting their application;
    - Once research is funded, DMPs should be updated annually by a designated member of the study team.
- It is the responsibility of the applicant to identify appropriate discipline-specific data repositories and then to deposit data in one of them to allow sharing. This should be done in a timely manner.
- A limited and defined period of exclusive data use is reasonable.
- The MRC is willing to cover some of the costs associated with data sharing.

# INTRODUCTION

In common with most other major funders of biomedical research, the Medical Research Council (MRC) expects all research projects it funds to manage effectively and, where possible, share research data. The MRC's overall concern is to "maximise the life-time value of research data assets for human health and to do so in a way that is timely, responsible, with as few restrictions as possible, and consistent with the law, regulations and recognised good practice".[2]

Effective management of medical research data can:

- enable new research questions to be answered using existing data;
- promote collaboration between different research teams and diverse disciplines;

[1] DMP template (.doc), https://mrc.ukri.org/documents/doc/data-management-plan-template/

[2] MRC Policy on Research Data Sharing, https://mrc.ukri.org/documents/pdf/mrc-data-sharing-policy/

- allow the sharing of knowledge about the best methods for data collection, linkage and analysis;
- help to ensure that collected data are clean and well documented, with value added;
- support the independent verification of established research findings;
- allow the development and testing of new research methods;
- help to use the data produced by study participants to best effect.

All applicants submitting funding proposals to the MRC are required to submit a Data Management Plan (DMP) as an attachment to the main Je-S Proposal form. This also applies to applications for the extension of current funding.

In your DMP you should demonstrate how you expect to fulfil your research data management responsibilities to the MRC (explained below), identify any obstacles to doing so and describe the measures you will take to meet these challenges. Your DMP can also help to justify any funding required to fulfil the MRC's data management requirements.

Where research is only partly funded by the MRC, the MRC research data requirements still apply. If these requirements conflict with the policy of another part-funder, you should discuss this with the MRC[3] before your application is finalised.

Your DMP will be reviewed as an integral part of the application; a poor DMP can have a negative impact on an otherwise strong application. However, the MRC believes that data sharers should receive "…full and appropriate recognition by funders, their academic institutions and new users for promoting secondary research".[4]

Applicants to The National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) are also required to provide a DMP,[5] containing the same information required by the MRC. Therefore, the issues discussed in this guide are equally relevant to both MRC and NC3Rs applicants.

## DMP format

It is highly recommended that applicants make use of the DMP template provided by the MRC.[6] NC3Rs also recommend the use of the MRC DMP the template. If applicants choose *not* to use the template, the points raised within the MRC template must still be addressed. If the funding application is successful, your DMP should be reviewed and updated annually by a designated member of your study team.

Your DMP should be written with two distinct audiences in mind: fellow scientists in your own field and data managers. Avoid generalisations and be concise. Before you write anything, consider the complexity of the data you intend to create and how

[3] MRC Head Office Research Programmes Group, MRCdatasharing@headoffice.mrc.ac.uk
[4] MRC Policy on Research Data Sharing, https://mrc.ukri.org/documents/pdf/mrc-data-sharing-policy/
[5] NC3RS policy on research data, http://www.nc3rs.org.uk/sites/default/files/documents/Funding/Handbook.pdf (p.28, 'f. Data Preservation for Sharing)
[6] MRC DMP template (.doc), https://mrc.ukri.org/documents/doc/data-management-plan-template/

much long-term value your data is likely to offer to the wider research community.

If the scale, complexity and cost of managing your data are low, your DMP could be less than half a page long. In the case of population cohorts, genetic and omics data, biobanks, and other datasets that are potentially rich resources for the wider community, your DMP should be between 1000 and 1500 words long (a maximum of 3 A4 pages).[7]

If your research will form part of an established population or patient-based study, your DMP will be significantly longer, as it must indicate how you will meet twenty-one additional requirements.[8] If this is the case, much of the information you will need to provide to the MRC should already be included in your study's existing Policy on Data Sharing. This should be available on your study's project website.

If you intend to establish a new population- or patient-based study, you should seek advice on how to work towards meeting the MRC's twenty-one data sharing requirements. In the first instance, contact the Research Data Service.[9]

## The nature of your research data

As part of your DMP, you should state the nature of your data (for example, qualitative, statistical, interview, or imaging) and in which format/s your data will be collected, analysed and stored (for example, Open Document Format, .CSV file or Excel spreadsheet). The key aim here is to explain how your research data will support not only your own immediate research needs but also future secondary analysis.

If you find you need to use a non-standard data format (for example for data from a unique, in-house system) which would be unsuitable for wider use, you should consider converting your data to a more widely used format once you are ready to share it. Explain this intention in your DMP.

If you're unsure which file formats to use, the UK Data Archive maintains a list of recommended deposit formats[10] which may be suitable.

You should also try to estimate the size of the data you expect to generate. This can be difficult to do before a study begins; if necessary, use quantities generated by similar past studies as a basis for your estimate.

## Re-using existing data

The MRC recognises that scientists use existing data in increasingly diverse ways: for instance, by using data linking or meta-analysis. Your Case for Support should identify existing datasets you expect to draw on. If you intend to generate new data, your Case for Support should also explain why this is necessary.

[7] MRC Guidance for applicants https://mrc.ukri.org/funding/guidance-for-applicants/2-the-application/#2.2.7

[8] MRC Policy on Sharing of Research Data From Population and Patient Studies, https://mrc.ukri.org/publications/browse/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies

[9] Research Data Service contacts, http://www.bristol.ac.uk/staff/researchers/data/contacts

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

Studies that make use of existing datasets should meet the same high standards as all MRC research in regard to scientific quality, ethical requirements and value for money. They should also add recognisable value to the original datasets. The MRC suggests that such research is often most fruitful when it is a collaboration between a new user and the original data creators.

If you have previously collected or generated research data of your own and you intend to use it as part of new research, you should ensure that your DMP describes procedures for managing both existing and newly generated data.

## Copyright

Unless stated otherwise, the ownership of intellectual property lies with the organisation carrying out the research. If you plan to work collaboratively on a new study with an external partner, copyright and other IPR (Intellectual Property Rights) issues may need to be clarified in a Consortium Agreement or Memorandum of Understanding. This is not required as part of your application, but you should say that if the application is successful, such an agreement will be created. All research partner institutions should then be made aware a Consortium Agreement will be forthcoming. Research Enterprise and Development[11] can provide advice on research IPR issues. Ultimately, however, it is the responsibility of the applicant to ensure that IPR issues do not unnecessarily prohibit data sharing.

## Ensuring the quality of your data

Your DMP should describe how you will ensure the quality of your research data. Quality should be considered whenever data is created or altered, for instance at the time of data collection or data entry. Procedures you may wish to carry out to ensure that data quality is maintained include: putting time aside to validate data manually, regular calibration, repeating samples, standardised data capture, or recording and entering values into prepared databases or transcription templates. You should mention in your DMP any data standards you intend to use at the data collection/generation stage (see Metadata, below).

## Data storage

You should explain where your data will be stored, how it will be organised in the short term, and who will back it up. If you are not part of a study with existing data storage arrangements, it is recommended that when you create data you store it in the University's Research Data Storage Facility (RDSF) managed by the Advanced Computing Research Centre (ACRC).[12] Each research staff member is entitled to 5TB of secure data storage without charge. If your storage quota is already used up, or if your project will exceed this storage limit, there will be a cost, and the ACRC should be contacted for guidance before your budget is finalised. The back-up procedures, policies and controlled access arrangements used by the RDSF are of a very high standard.

If you do not intend to make use of the RDSF, your storage provider's back-up procedures should be

[11] Research Enterprise and Development, University of Bristol, http://www.bristol.ac.uk/red/contracts/

[12] Advanced Computing Research Centre, University of Bristol, www.acrc.bris.ac.uk

described instead. If you will be working collaboratively with other institutions, make sure that the security and back-up procedures of each data-holding partner are described in your DMP.

Your DMP should also outline how you will keep your data safe *before* it's deposited in a storage facility such as the RDSF. This is particularly important if you are conducting field research. As a minimum requirement, always try to ensure that at least two copies of the data exist and that every copy can easily be accounted for and located if required.

## Potentially disclosive personal information

If your research is part of an established study, data security measures will already be in place and these should be summarised in your DMP. However, all MRC-funded researchers have a responsibility to ensure that opportunities for data re-use are maximised, within the regulatory requirements of the law. MRC adopts the view that the potential benefits to patients and the public should outweigh identified risks. The MRC's guide to using Personal Information in Medical Research states:

*"Principal investigators must take personal responsibility for ensuring (as far as is reasonably practical) that training, procedures, supervision, and data security arrangements are sufficient to prevent unauthorised breaches of confidentiality."* [13]

If you are planning to make use of a data storage facility that complies with a recognised information security standard (such as ISO 27001), explain this in your DMP. If not, describe the main risks to the security of any data relating to human participants and how these risks will be addressed, such as by access control or encryption. The information you provide must correspond to that provided in the plan made as part of your ethical review.

## Metadata

Metadata is 'data about data' or 'cataloguing information' that enables data users to find 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. It is generally best to use established and shared metadata standards, rather than create new ones. This helps with consistency and saves effort.

For example, ICD-10[14] (International Statistical Classification of Diseases and Related Health, 10th Revision) is a medical classification from the World Health Organisation that "codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases". ICD-10 provides 14,400 unambiguous codes.

The Data Documentation Initiative[15] (DDI) is an international standard for describing data from the social, behavioural, and economic sciences. DDI uses XML to allow metadata to be encoded in a standardised way, simplifying data sharing and subsequent re-use.

[13] Personal Information in Medical Research, https://mrc.ukri.org/documents/pdf/personal-information-in-medical-research/

[14] International Classification of Diseases ICD-10, version:2016, icd.who.int/browse10/2016/en

[15] Data Documentation Initiative, www.ddialliance.org/getting-started

Metadata can be kept in a separate, dedicated database or spreadsheet. If you are planning to use data analysis software, such as a qualitative analysis package, you will have the option of adding documentation within the software itself, in the form of notes.

In attempting to organise and document your data, it may help to imagine a secondary data user trying to make sense of your data in your absence, after your project has concluded. If no metadata were provided, this secondary user would be faced with the difficult task of 'unpicking' your data. How, for instance, would they make sense of your file and folder names? Or your methodology or approach to data processing? What extra information would they need to make the most of your data?

## Providing long term access to data

If your research is part of an established study, the existing Study Policy on Data Sharing should contain much of the information you need to provide about current data governance procedures and existing data-sharing agreements. If your study has an established history of data sharing, your DMP should include a brief evaluation of these activities.

If there is not an established history of data sharing, it is the responsibility of the applicant to identify appropriate discipline-specific data repositories. The Wellcome Trust maintains a list of major data repositories including biomedical repositories that preserve and provide access to research data.[16]

Researchers may choose to share their data by depositing it in a repository such as the UK Data Archive.[17]

The University of Bristol has its own Research Data Repository providing several different levels of access to data which researchers from any discipline may wish to use. Access options range from entirely open to rigorously controlled, which is suited to 'sensitive' data. 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[18] as early as possible if you believe you'll need to make use of Bristol's data repository.

Data must be shared in a timely and responsible manner. However, the MRC recognises that ongoing research contributing to the completion of datasets must not be compromised by premature sharing and analysis, therefore a limited and defined period of exclusive data use is reasonable. Access to data may also be delayed for a short period to allow time for the preparation and filing of patent applications. You should indicate which data cannot be retained and/or shared, and explain why this is necessary.

## Roles and responsibilities

Data management responsibilities should be clearly assigned to named individuals within your DMP. The

[16] Wellcome Open Research approved data repositories https://wellcomeopenresearch.org/for-authors/data-guidelines#hosting

[17] UK Data Archive www.data-archive.ac.uk

[18] The University of Bristol's Research Data service *data.bris*, https://data.bris.ac.uk/contact/

Principal Investigator (or Research Director for larger studies) is usually responsible for research data sharing. The person responsible for maintaining and updating the DMP (if this is different) should also be named in your DMP.

Several support services are in place at the University of Bristol to help you manage and share your research data, and any such services you plan to use should be mentioned in your DMP. These services include: ACRC (secure data storage), your Zonal IT team (general IT support), the data.bris service (research data management training and guidance), RED (advice on collaborative projects and IPR) and the Office of the Secretary (for Data Protection and FOI).

### The cost of managing research data

If any costs are involved in meeting the MRC's data management requirements (for example, the cost of dedicated effort, equipment or software tools for managing, storing or providing access to your data), these should be mentioned in your application.

If the costs are substantial, you should differentiate between:

- costs associated with collecting and/or processing new data;
- your own research on newly acquired and legacy data;
- ongoing data curation and preservation;
- providing access and data sharing.

## CITING RESEARCH DATA IN RESEARCH OUTPUTS

From 1st April 2013 all the UK's Research Funding Councils, as part of RCUK (now UKRI), 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. Or the enquirer might be directed to a page which displays 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 publication, 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 will change over time. If you plan to use an established data repository service, ask it for a unique reference identifier which could be included in the publication instead. If you're not planning to use an established data repository, contact the Research Data Service[19] for further guidance.

[19] The University of Bristol's Research Data Service http://data.bris.ac.uk/contact/

# SAMPLE MRC DATA MANAGEMENT PLAN

## 0. Proposal name

Investigating the effects of therapeutic agents on stem cell populations in childhood leukaemia.

## 1. Description of the data

### 1.1 Type of study

In this study we will investigate the effects of therapeutic agents on cells from children with acute lymphoblastic leukaemia. We will also characterise the cells that may be responsible for initiating and maintaining this malignancy.

### 1.2 Types of data

Quantitative data derived from our experimental approaches and statistical analyses of these data will be managed along with Genotypic data from results of our microarray studies. There will also be administrative data on linked anonymised tissue samples and storage locations of the samples.

### 1.3 Format and scale of the data

Data will initially be collected in a variety of file formats, mainly Microsoft Excel for databases, MS Word and equipment specific software such as FlowJo for flow cytometry data. Graph Pad prism will be used to generate graphs and conduct statistical analyses. Presentations will usually be compiled using MS Powerpoint. These files will also be stored in Open Document Format or as .CSV files, for spreadsheets, to enable sharing. Digital images will be saved as .TIFF files for this purpose.

## 2. Data collection / generation

### 2.1 Methodologies for data collection / generation

Data will be generated from the outlined experimental protocols, mainly by flow cytometric analysis, using FlowJo software. FlowJo permits analyses of such data and generation of figures. These can be exported directly to Excel, Powerpoint or saved in acceptable formats for sharing and re-use, as defined by the UK Data Archive. The information generated in the planned experiments will be added to our existing data sets which contain information of the ability of particular patient samples to engraft, the response of these samples to therapeutic agents, the frequency of stem cell populations in particular samples. This will enable us to build a large dataset of information, which will be invaluable for the current grant proposal, future proposals and for work with collaborators.

### 2.2 Data quality and standards

Such standards will be met through using equipment, which is calibrated, at regular intervals in accordance with good laboratory practice e.g. daily calibration of flow cytometers and incubators to minimise variation due to equipment. Equipment calibration is logged and monitored using Q-pulse document control system. Data capture will be standardised using appropriate software e.g. FlowJo can directly export data generated in other file formats, thus avoiding errors through manual input of data. Where manual input is the only option, consistency will be maintained through peer review and cross checking the results with existing measurements.

## 3. Data management, documentation and curation

### 3.1 Managing, storing and curating data.

All data will be backed up immediately after generation onto the University of Bristol Research Data Storage Facility (RDSF). The RDSF provides secure, long-term storage for research data. This two million pound investment provides nightly backup of all data, with further resilience provided by three geographically distinct storage locations. A tape library is used for backup purposes and also for long-term, offline data storage. Only authorised users can access data stored within the RDSF.

The RDSF is managed by Bristol's Advanced Computing Research Centre (ACRC) which has a dedicated steering group and a rigorous data storage policy (https://www.acrc.bris.ac.uk/acrc/RDSF_policy.pdf). The RDSF upholds and reinforces Bristol's wider Information Security Policy (http://www.bristol.ac.uk/media-library/sites/infosec/documents/ISP-01v1.2.pdf).

### 3.2 Metadata standards and data documentation

Methods used to generate the data will be described as standard operating procedures in an open file format. Metadata information on sample ID, instrument settings, experimental variables, operator ID are already captured for every sample run on the flow cytometer and can be easily exported to open file format.

### 3.3 Data preservation strategy and standards

Data will be stored in the RDSF for at least 20 years. Any data selected for publication in the Research Data Repository (see below) will also be publicly available for 20 years.

## 4. Data security and confidentiality of potentially disclosive personal information

### 4.1 Formal information/data security standards

Our Ethical review allows collection of linked anonymised data from patient samples. No personal data will be included in our data and therefore there is no risk of disclosure of personal information.

## 5. Data sharing and access

### 5.1 Suitability for sharing

Yes.

### 5.2 Discovery by potential users of the research data

New users can find out about our data through CLR-UK meetings and through the publication of our datasets in the University of Bristol Research Data Repository (data.bris). The data.bris Research Data Repository offers a means for Bristol's researchers to openly share non-confidential research data, without the need for external data users to undergo any form of authentication. Each deposit is accompanied by appropriate metadata and is assigned a unique Digital Object Identifier (DOI) via the DataCite scheme, allowing it to be cited in publications.

### 5.3 Governance of access

Data will be publicly available through the data.bris repository, published under a permissive re-use license.

### 5.4 The study team's exclusive use of the data

Our intended policy is that the team should have exclusive use of the data for a period of 12 months or until the data is published or patent applications have been filed. Data will be shared with named collaborators during this time.

### 5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

Delays to the above data sharing policy may only arise through IPR. We will seek advice from the University Research and Enterprise Development Office if we think an output is worthy of registering as IPR. This consultation will take place prior to any publication or disclosure of results.

### 5.6 Regulation of responsibilities of users

N/A

## 6. Responsibilities

The PI and the senior members of their laboratory will be responsible for verifying data is accurate and records are up to date. Staff from the Advanced Computing Research Centre, Research Enterprise and Development and the Research Data Service will assist with data storage, curation and sharing.

## 7. Relevant institutional, departmental or study policies on data sharing and data security

*Please complete, where such policies are (i) relevant to your study, and (ii) are in the public domain, e.g. accessibly through the internet.*

*Add any others that are relevant*

| Policy | URL or Reference |
|---|---|
| Data Management Policy & Procedures | http://www.bristol.ac.uk/research/environment/governance/research-data-policy/ |
| Data Security Policy | https://www.acrc.bris.ac.uk/acrc/RDSF_policy.pdf |
| Data Sharing Policy | http://www.nationalarchives.gov.uk/doc/non-commercial-government-licence/version/2/ |
| Institutional Information Policy | http://www.bristol.ac.uk/media-library/sites/infosec/documents/ISP-01v1.2.pdf |
| Other: | |
| Other | |

## 8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their telephone & email contact details