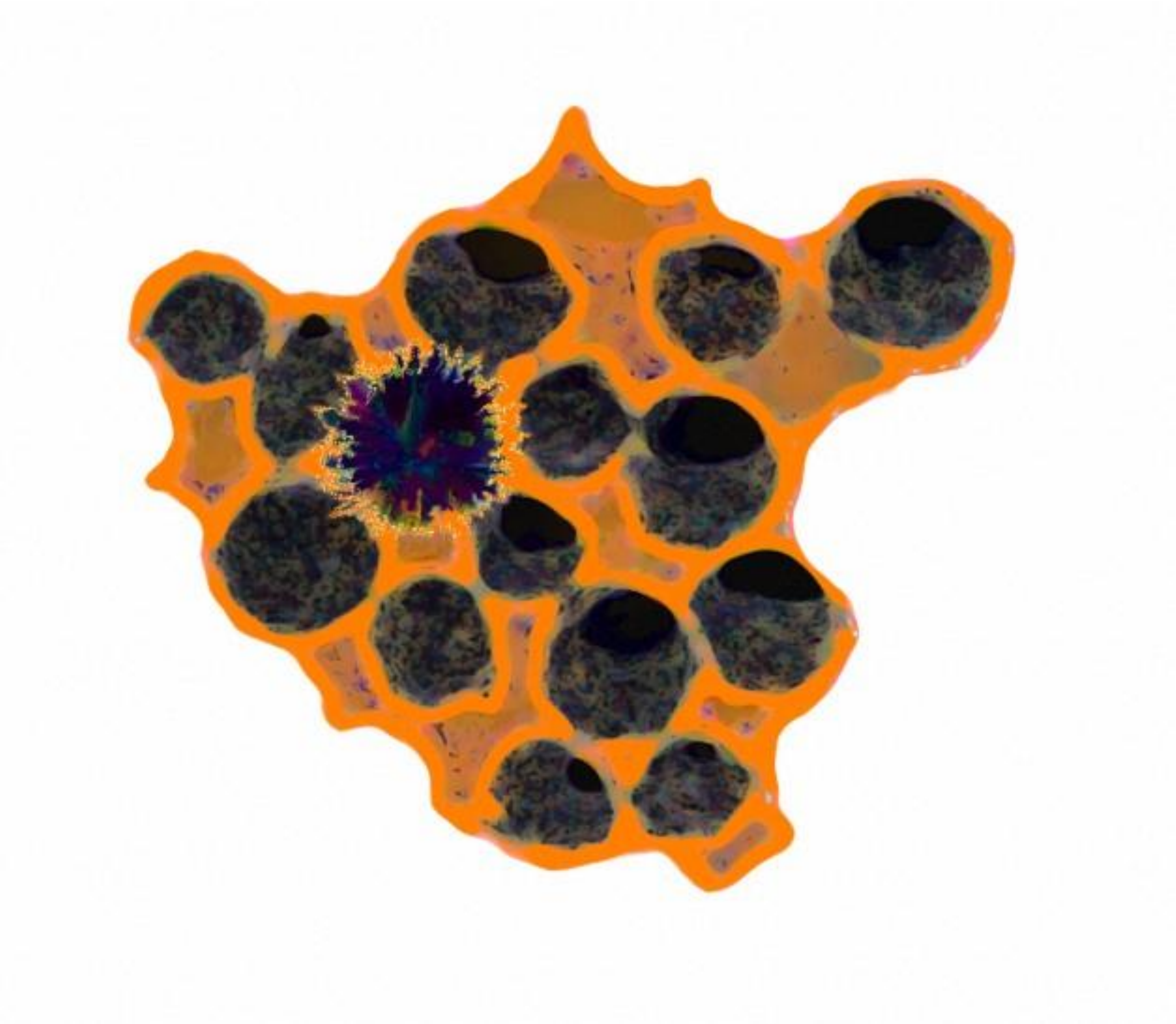
## Data Management Planning

# CRUK funding applicants

Version 2.3 June 2021



## University of Bristol

## Research Data Service

Image: Abnormal cell – artwork, Benita Denny,

Wellcome Images, CC-BY-NC-ND 4.0

# SUMMARY

- Data should be made available with as few restrictions as possible, whilst respecting confidentiality, commercial agreements and intellectual property rights (IPR);
- A limited period of exclusive data use is acceptable;
- Data sharing does not alter CRUK's support of IPR to maximise benefit to patients;
- Applicants should produce a Data Management and Sharing Plan as part of the application process;
- The Plan will be reviewed as part of the funding decision;
- Funding committees will monitor the progress of implementation of the Plan, whilst accepting changes may need to be made as the project progresses;
- No set template is provided for the Data Management and Sharing Plan, but it should cover eight key areas;
- It is accepted that the methods for sharing will vary based on the types of data produced;
- Data should be preserved for a minimum of five years after the end of a project;
- CRUK are willing to cover justified costs associated with research data management.

# INTRODUCTION

In common with most other major funders, Cancer Research UK expects all data generated through its funding to be considered for sharing and made as widely and freely available as possible. It is committed to ensuring that "the data generated through its funding should be put to maximum use by the cancer research community and, whenever possible, is translated to deliver patient benefit".[1]

The CRUK Data Sharing and Preservation Policy[2] applies to all candidates seeking funding after 1st April 2009 and focuses on:

- The sharing of final research data for research purposes;
- Basic research, clinical studies, surveys and other types of research supported by CRUK;
- Unique data that cannot be replicated;
- Projects that transform or link pre-existing datasets.

Data from all activities relating to Phase I and Phase II clinical trials sponsored by CRUK are not automatically covered by the Data Sharing and Preservation Policy. Studies should contact the Centre for Drug Development on a trial-by-trial basis for further clarification.

A Data Management and Sharing Plan is required from all applicants seeking funding from CRUK as part of their grant proposal. If applicants feel data sharing is not appropriate, they must provide a clear explanation why not. The plan will be reviewed as part of the funding decision, with funding committees assessing

---

[1] CRUK Data sharing guidelines, http://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines

[2] CRUK Data Sharing Policy, https://www.cancerresearchuk.org/sites/default/files/cruk_data_sharing_policy_2020_final.pdf

the suitability of the plan, providing specific feedback where necessary, or even requesting revisions before a grant award letter is issued. Funding committees will then monitor progress in implementing Data Management and Sharing Plans, though it is accepted that methods and timelines for sharing data may need to be adapted during the course of a study.

This guide builds on information taken both from CRUK's 'Data Sharing Guidelines'[3] and 'Data Sharing FAQs'[4] which provide more detailed guidance on their expectations and requirements, and, in the case of sharing code, from email communication from CRUK to the University of Bristol's Research Data Service.

# DATA MANAGEMENT AND SHARING PLAN

As data sharing strategies will vary according to the type of data collected, CRUK do not specify the exact content or format of the plan. Depending on the funding committee, a box for completing a Data Management and Sharing Plan is either incorporated into the grant application form or provided as a separate document. The Science Committee and Clinical Research Committee expect a short, free-text description of how applicants plan to adhere to the CRUK's policy at the grant application stage. A more detailed Data Management and Sharing Plan, in consultation with CRUK representatives, will then be requested if an application is successful. The Population Research Committee expect a full Data Management and Sharing Plan at the application stage.

The DMPonline tool provides a template for CRUK data management and sharing plans, with sections covering all CRUK's requirements.[5] University of Bristol researchers can register for the tool using their University sign in.

The following areas should be considered when producing a CRUK Data Management and Sharing Plan:

- The volume, type, content and format of the final dataset;
- The standards that will be utilised for data collection and management;
- The metadata, documentation or other supporting material that should accompany the data for it to be interpreted correctly;
- The method used to share data;
- The timescale for public release of data;
- The long-term preservation plan for the dataset;
- Whether a data sharing agreement will be required;
- Any reasons why there may be restrictions on data sharing, for example;
    - Development arrangements through Cancer Research Technology including intellectual property protection and commercialisation;

---

[3] CRUK Data Sharing Guidelines, http://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines

[4] CRUK Data Sharing FAQs, http://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-faqs

[5] DMPonline, https://dmponline.dcc.ac.uk/

- Proprietary Data – restrictions due to collaborations with for profit organisations;
- International policies governing the sharing of data collected outside of the UK;
- Confidentiality, ethical or consent issues that may arise with the use of data involving human subjects.

## Data types, formats and volumes

As part of your Data Management and Sharing Plan, you should state the types of data you will be producing (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 plan. If you're unsure which file formats to use, the UK Data Archive maintains a list of recommended deposit formats[6] 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.

## Standards and data quality

Your plan 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 plan any data standards you intend to use at the data collection/generation stage (see Metadata, below).

## Metadata and documentation

Metadata is 'data about data' or 'cataloguing information' that enables data users to find or use a dataset. In your Data Management and Sharing Plan you should outline how you propose to document your research data to meet both your own needs and those of later users. CRUK expect this documentation to include such information as the methodology used to collect data, definitions of variables, units of measurement, any assumptions made, the format of the data, file type of the data etc. To support this, researchers are strongly encouraged to utilise community standards to describe and structure data, (e.g. common terminology, minimum information guidelines and standard data exchange formats),

---

[6] UK Data Archive File Formats Table, https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats

rather than create new ones. This helps with consistency and saves effort. The FAIRsharing[7] portal has useful lists of discipline-specific metadata schemas.

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?

## Data sharing

CRUK expect all data to be considered for sharing, and to be made as widely available as possible, whilst respecting confidentiality, commercial agreements and intellectual property. CRUK have communicated to us by email that if sharing code together with project data will facilitate the reuse of that data - especially of data that underpins findings in publications – then CRUK recommend sharing of the code. CRUK also suggest that if studies are willing and see value to the scientific community in sharing further code then they are encouraged to do so.

Data sharing should occur in a timely manner. It is acknowledged that researchers should be allowed to benefit from the data they have generated, and investigators are allowed a period of private use of their data, but not prolonged exclusive use.

Data is expected to be released no later than the acceptance of publication of findings from the dataset, or in line with any procedures of the relevant research area (for example crystallography data). This is unless any restrictions from IP or third-party agreements still apply. For experiments conducted over a prolonged period (e.g. population studies) it is expected that subsets of the data will be made available for sharing, whilst the researcher can continue to benefit from a period of exclusive analysis of the dataset as a whole.

CRUK acknowledge that the methods for sharing will vary based on the types of data produced. Data sharing can be done by any of the following methods:

*Under the auspices of the Principal Investigator* – if using this method, investigators may securely send data to a requestor, or upload to their institutional website. If a PI chooses to control access to data themselves, CRUK recommend the use of a Data Access Agreement. This will ensure responsibilities of both parties, along with other rights, are agreed at the outset. CRUK recommend referring to NCRI's 'Samples

---

[7] FAIRsharing https://fairsharing.org/

and Data for Research: Template for Access Policy Development' for cancer for more guidance.[8]

Sharing data through a PI does have implications; providing contact details or a URL for underlying data in a published article may not be acceptable to publishers, many of whom now prefer a Digital Object Identifier (DOI) to be used. There is also potentially a considerable amount of administrative work in managing and monitoring access requests. Researchers should therefore think carefully before choosing this option as the only way of providing access to data.

*Through a third party* – investigators can transfer data to a data archive where it will be made more widely available to the scientific community. CRUK state that data archives or repositories are particularly suitable for those who will be potentially handling large volumes of requests for data (especially if these need to be vetted), or if technical assistance is required to help secondary users to analyse the datasets. Most data archives or repositories now provide a Digital Object Identifier (DOI) for published datasets to allow them to be easily cited in research publications.

There are an increasing number of discipline-specific data repositories available. The Wellcome Trust maintains a list of major data repositories including biomedical repositories that preserve and provide access to research data.[9] Researchers may choose to share their data by depositing it in a repository such as the UK Data Archive.[10]

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. 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[11] as early as possible if you believe you'll need to make use of Bristol's data repository.

*Using a data enclave* – in some instances, datasets which cannot be made publicly available due to confidentially issues or third-party licensing restrictions, may be accessed through a data enclave. This provides a controlled secure environment where approved researchers can perform analyses using restricted data resources.

It is acceptable to use a combination of these methods, for example if working with different versions or varying access control requirements.

---

[8] https://www.ncri.org.uk/wp-content/uploads/Initiatives-Biobanking-2-Access-template-1.pdf

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

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

[11] The University of Bristol's Research Data service *data.bris*, http://www.bristol.ac.uk/staff/researchers/data/contacts/

## Restrictions

Not all data is suitable for open sharing, and CRUK acknowledge that the following restrictions may exist:

*Intellectual Property Rights* – any IP issues or plans should be outlined in the data sharing plan. CRUK understand that some research, particularly that with a translational focus, are likely to result in patents or commercial collaborations, and this should be discussed with your technology transfer office and Cancer Research Technology prior to data sharing. The filing of patents is encouraged, but whilst a subsequent delay in the release of data may be necessary, it should not hinder data sharing.

*Commercial agreements/proprietary data* – Any issues around data sharing as a result of private sector co-funding should be outlined in the plan. Alternative ways that data requests may be considered should be explored by the applicant.

*Research involving human participants* – investigators must ensure appropriate consent is gained to share data, alongside ethical approval. Data should be anonymised prior to sharing, and any indirect identifiers that may lead to disclosures should be removed. If data cannot be fully anonymised, or risk will still remain, this needs to be outlined in the Data Sharing Plan. The Research Data Service has produced a guide to sharing data involving human participants, which includes sample statements for consent forms.[12]

## Data storage and preservation

You should explain where your data will be stored, how it will be organised, who will back it up, and how it will be preserved for the long-term. 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).[13] 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 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 plan.

Your Data Management and Sharing Plan 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, try to ensure that at all times at least two copies of the data exist, and that every copy can easily be accounted for and located if required.

---

[12] Sharing Research Data Concerning Human Participants, http://bit.ly/35hldfU

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

You should state in your plan how your data will be preserved beyond the life of the project. CRUK expect researchers to preserve all data resulting from a grant so that it can be used for new or follow up studies. Data is expected to be preserved and available for sharing for a minimum of five years following the end of a research grant. The RDSF provides secure storage for a minimum of ten years.

## The cost of managing research data

CRUK regard the management and sharing of research data as a fundamental aspect of good scientific practice and therefore will fund justified running costs associated with data management and sharing activities. You should include any expected costs in your application and if these 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.