

## **CRUK Template**

**Plan Name** Standard CRUK Template

**Principal Investigator / Researcher** TUoS Researcher

**Funder** -

**Institution** University of Sheffield

### **Data description**

**Outline the volume, type, content and format of the final dataset**

#### **DCC guidance on Data Volumes**

Questions to consider:

- Do you have sufficient storage?
- Do you need to include costs for additional managed storage?
- Will the scale of the data pose challenges when sharing or transferring data between sites?

Guidance:

Consider the implications of data volumes in terms of storage, backup and access. Estimate the volume of data in MB/GB/TB and how this will grow to make sure any additional storage and technical support required can be provided.

#### **The University of Sheffield: guidance on Data Volumes**

Please see the University of Sheffield Corporate Information and Computing Services webpages on '[Research data storage](#)' and '[Storage options](#)' for guidance.

#### **DCC guidance on Data Type**

Questions to consider:

- What types of data will you create?
- Which types of data will have long-term value?

Guidance:

Outline the types of data that are expected to be produced from the project e.g. quantitative, qualitative, survey data, experimental measurements, models, images, audiovisual data, samples etc. Include the raw data arising directly from the research, the reduced data derived from it, and published data.

#### **The University of Sheffield: guidance on Data Type**

Please see the University of Sheffield webpage '[What is research data?](#)' for guidance.

#### **DCC guidance on Data Format**

Questions to consider:

- What format will your data be in?
- Why have you chosen to use particular formats?
- Do the chosen formats and software enable sharing and long-term validity of data?

Guidance:

Outline and justify your choice of format e.g. SPSS, Open Document Format, tab-delimited format, MS Excel. Decisions may be based on staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community. Using standardised and interchangeable or open lossless data formats ensures the long-term usability of data.

See UKDS Guidance on [recommended formats](#).

### **The University of Sheffield: guidance on Data Format**

Please see the University of Sheffield webpage on '[Organising your data: Choosing data formats](#)' for guidance.

### **Data standards**

**State the standards that will be utilised for data collection and management**

### **DCC guidance on Data Capture Methods**

Questions to consider:

- How will the data be created?
- What standards or methodologies will you use?
- How will you structure and name your folders and files?
- How will you ensure that different versions of a dataset are easily identifiable?

Guidance:

Outline how the data will be collected/generated and which community data standards (if any) will be used at this stage. Indicate how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier for the research team to find, understand and reuse.

### **The University of Sheffield: guidance on Data Capture Methods**

How will you organise your research data and handle version control? Please see the University of Sheffield webpage on '[Organising your data: Naming and organising files and folders](#)' for guidance.

### **DCC guidance on Data Quality**

Questions to consider:

- How will you control data capture to ensure data quality?
- What quality assurance processes will you adopt?

Guidance:

Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.

### **The University of Sheffield: guidance on Data Quality**

Please see the UK Data Service webpage on [data quality assurance](#) for guidance.

### **Metadata and documentation**

**Outline the metadata, documentation or other supporting material that should accompany the data for it to be interpreted correctly**

#### **CRUK Guidance**

For data sharing to be a success it is important that data are prepared in such a way that those using the dataset have a clear understanding of what the data mean so that they can be used appropriately. To enable this, applicants are encouraged to include with the dataset all the necessary information (metadata) describing the data and their format. This information should 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).

### **DCC guidance on Metadata**

Questions to consider:

- How will you capture / create the metadata?
- Can any of this information be created automatically?
- What metadata standards will you use and why?

Guidance:

Metadata should be created to describe the data and aid discovery. Consider how you will capture this information and where it will be recorded e.g. in a database with links to each item, in a 'readme' text file, in file headers etc.

Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. The DCC offers a [catalogue of disciplinary metadata standards](#).

### **The University of Sheffield: guidance on Metadata**

Metadata is a structured form of documentation that identifies and describes your data. Researchers should use community standards, where they exist: see the DCC webpage on [Disciplinary metadata standards](#).

Please see the University of Sheffield webpages on '[Describing your data](#)' and '[Metadata](#)' for guidance.

## DCC guidance on Documentation

Questions to consider:

- What metadata, documentation or other supporting material should accompany the data for it to be interpreted correctly?
- What information needs to be retained to enable the data to be read and interpreted in the future?

Guidance:

Describe the types of documentation that will accompany the data to provide secondary users with any necessary details to prevent misuse, misinterpretation or confusion. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data.

## The University of Sheffield: guidance on Documentation

**Note:** Documentation and metadata describe the context, content and structure of your data and are essential for understanding and reusing them. See the University of Sheffield webpage '[Describing your data](#)' for more information.

**Example text:** *“Methods and SOPs will be stored electronically in Microsoft Word documents (.doc) with the spreadsheets containing data”*

*“Explanation of the experimental and analytical methods used will be provided in text documents, stored alongside the data”*

*“Data documentation will accompany datasets submitted to the ... repository at the end of the research”*

## Data sharing

### Outline the method used to share data

#### CRUK Guidance

The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Data can be shared by any of the following methods:

- **Under the auspices of the Principal Investigator** - Investigators sharing under their own auspices may securely send data to a requestor, or upload the data to their institutional website. Investigators should consider using a data-sharing agreement (see below) to impose appropriate limitations on the secondary use of

the data.

- **Through a third party** - Investigators can share their data by transferring it to a data archive facility to distribute more widely to the scientific community, to maintain documentation and meet reporting requirements. Data archives are particularly attractive for investigators concerned about managing a large volume of requests for data, vetting frivolous or inappropriate requests, or providing technical assistance for users seeking to help with analyses.
- **Using a data enclave** - Datasets that cannot be distributed to the general public due to confidentiality concerns, or third-party licensing or use agreements that prohibit redistribution, can be accessed through a data enclave. A data enclave provides a controlled secure environment in which eligible researchers can perform analyses using restricted data resources.
- **Through a combination of methods** - Investigators may wish to share their data by a combination of the above methods or in different versions, in order to control the level of access permitted.

## DCC guidance on Method For Data Sharing

Questions to consider:

- How will you make the data available to others?
- With whom will you share the data, and under what conditions?

Guidance:

Consider where, how, and to whom the data should be made available. Will you share data via a data repository, handle data requests directly or use another mechanism?

The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Mention earlier examples to show a track record of effective data sharing.

## The University of Sheffield: guidance on Method For Data Sharing

**Note:** At the end of your research project, your funder may require you to make your research data available for sharing with as few restrictions as possible. Data may be shared by being published in:-

- a Repository or Data Centre - see the University of Sheffield webpage on '[Research data repositories](#)' for guidance
- a journal as an article's supplementary material
- a data journal as a data paper.

Wherever data is published, a metadata record should be [registered in ORDA](#), the University of Sheffield data repository.

**Suggested text for use when data will be placed in a repository:** *“Data will be made available through shared research platforms [insert repository / platform relevant to project] with the relevant permissions in place.”*

**Suggested text for use when data will not be placed in a repository:** *“The lead PI and project team [including collaborators if applicable] will review applications to access experimental data and make the decision on whether to supply research data to potential applicants. Data will then be released on a case by case basis.”*

## **DCC guidance on Data Repository**

Questions to consider:

- Where (i.e. in which repository) will the data be deposited?

Guidance:

Most research funders recommend the use of established data repositories, community databases and related initiatives to aid data preservation, sharing and reuse.

An international list of data repositories is available via [Databib](#) or [Re3data](#).

## **The University of Sheffield: guidance on Data Repository**

**Note:** For guidance see the University of Sheffield webpages on '[Publishing and sharing your research data](#)' and '[Data repositories](#)'.

Long term preservation and access may be best managed by using a specialist data repository. Your funder may specify a data repository to use, such as [UK Data Service ReShare](#), [NERC Data Centres](#) or [Archaeology Data Service](#). Alternatively, look in [re3data.org](#), at [Wellcome Trust - Data repositories and database resources](#) and at [BBSRC supported resources](#) to find an appropriate repository. If no suitable repository is available you may [deposit data in ORDA](#), the University of Sheffield data repository. Alternatively, if you need to regulate users' access through 'Data sharing agreements', data may be retained in the University's research storage infrastructure and [registered in ORDA](#).

## **State the timescale for public release of data**

### **CRUK Guidance**

As the value of data is often dependent on its timeliness Cancer Research UK expects that data sharing should occur in a timely manner. Cancer Research UK acknowledges that the investigators who generated the data have a legitimate interest in benefiting from their investment of time and effort and we therefore support the initial investigator having a reasonable period of private use of the data but not prolonged exclusive use.

Cancer Research UK expects data to be released no later than the acceptance for publication of the main findings from the final dataset (unless restrictions from third party agreements or IP protection still apply) or on a timescale in line with the

procedures of the relevant research area. For example, for crystallography data there is an agreed 12-month delay between publishing the first paper on a structure and making the co-ordinates public.

With experiments carried out over an extended period of time, (e.g. population based studies), it is reasonable to expect that subsets of data analysed by the investigator(s) be made available for sharing. The investigator(s) can then continue to benefit from further reasonable periods of exclusive analysis while the dataset as a whole matures.

## **DCC guidance on Timeframe For Data Sharing**

Questions to consider:

- When will you make the data available?

Guidance:

Data (with accompanying metadata) should be shared in a timely fashion. It is generally expected that timely release would be no later than publication of the main findings and should be in-line with established best practice in the field. Researchers have a legitimate interest in benefiting from their investment of time and effort in producing data, but not in prolonged exclusive use. Research funders typically allow embargoes in line with practice in the field, but expect these to be outlined up-front and justified.

## **The University of Sheffield: guidance on Timeframe For Data Sharing**

**Note:** At the end of your research project, your funder may require you to make your research data available for sharing with as few restrictions as possible. Most funders allow a delayed release to allow researchers to have exclusive use of their data and to exploit the results of their research. See the University of Sheffield '[Research funder policy summaries](#)' webpage to determine when you need to make your data available.

**Suggested text in all cases:** *“The project group (including collaborators) will have exclusive use of the data until the main research findings are published or patent applications have been filed [if potentially relevant to project]” and/or “...or for a period of x months/years.”*

**Suggested text if delays are foreseen:** *“Delays in sharing data may arise through a delayed ability to analyse or publish the research findings.” and/or “Delays in sharing data may arise due to IPR and if this is a factor, advice will be sought from the University’s Research & Innovation Services.”*

**Optional additional text:** *“Following publication, data will be made available on request or shared through the [relevant research platforms].”*

## **Explain any reasons why there may be restrictions on data sharing**

### **CRUK Guidance**

Data which might have the potential to be exploited commercially or otherwise to deliver patient benefit should be discussed with your technology transfer office and

Cancer Research Technology prior to data sharing. Cancer Research UK encourages the appropriate filing of patents and recognises that there may be a need to delay the release of data until patent applications have been filed. Whilst there may be a delay in the release of data due to the application process, appropriate intellectual property protection should not hinder data sharing and may be the best way of ensuring that patient (and public) benefit is delivered. Any intellectual property issues or plans for commercialisation that may affect data sharing should be addressed in the data sharing plan. Cancer Research UK understands that unexpected intellectual property may arise during the course of the study and investigators may need to depart from their data sharing plan to protect intellectual property and for any other necessary steps to be taken. Data sharing may also be affected when co-funding is provided by the private sector (e.g. by a pharmaceutical company) or host institution resulting in some restrictions on the disclosure of data. For example with clinical trials, the Trial Management Group and/or trial sponsor etc may impose restrictions on data access. Any restrictions should be outlined in the data sharing plan and applicants should explore ways data sharing requests can be considered by the body that owns the data.

e.g. Development arrangements through Cancer Research Technology including intellectual property protection and commercialisation

e.g. Proprietary Data - restrictions due to collaborations with for profit organisations  
International policies governing the sharing of data collected outside of the UK

My research seeks supports from both the public and private sectors. How do I deal with the sharing of data? Where research is funded by a commercial sponsor, restrictions on data sharing may apply in arrangements agreed with the sponsor. Any such restriction(s) should be highlighted in the data management and sharing plan. In the event that researchers apply for or receive commercial funding for any part of their research that Cancer Research UK supports they should advise Cancer Research Technology of the situation without delay.

e.g. Confidentiality, ethical or consent issues that may arise with the use of data involving human subjects.

Investigators carrying out research involving human participants must ensure that consent is obtained to share information; furthermore the necessary legal, ethical and regulatory permissions regarding data sharing should be in place prior to disclosing any data. Every effort must be made to protect the identity of participants and, prior to sharing, data should be anonymised. In addition, any indirect identifiers that may lead to deductive disclosures should be removed to reduce the risk of identification. In most instances, sharing data should be possible without compromising the confidentiality of participants but if there are circumstances where data needs to be restricted due to the inability to protect confidentiality this should be fully addressed in the data management and sharing plan.

## **DCC guidance on Restrictions on Sharing**

Questions to consider:

- Are any restrictions on data sharing required? e.g. limits on who can use the data, when and for what purpose.
- What restrictions are needed and why?
- What action will you take to overcome or minimise restrictions?

Guidance:

Outline any expected difficulties in data sharing, along with causes and possible measures to overcome these. Restrictions to data sharing may be due to participant confidentiality, consent agreements or IPR. Strategies to limit restrictions may include: anonymising or aggregating data; gaining participant consent for data sharing; gaining copyright permissions; and agreeing a limited embargo period.

### **The University of Sheffield: guidance on Restrictions on Sharing**

**Note:** At the end of your research project, your funder may require you to make your research data available for sharing with as few restrictions as possible. Restrictions on the release of data may be allowed, to protect confidentiality and for other ethical and legal considerations:-

- Does your data include confidential and sensitive information?
- Have participants given consent for their data being shared?
- Consider what can be done to make sensitive data openly sharable - can these data be anonymised?
- If different parts of your research data require different access conditions, separate them and deposit them separately, applying different access conditions.

See the University of Sheffield '[Regulatory requirements](#)' webpage and the UK Data Service '[Legal and ethical issues](#)' webpages for more information.

**Suggested text if no restrictions are foreseen:** *“At present we do not foresee any delays in data sharing following publication of the main research findings.”*

**Suggested text for patient-based studies:** *“Patients will be made aware of our data sharing procedures at the time of consent.”*

### **DCC guidance on IPR Ownership and Licencing**

Questions to consider:

- Who owns the data?
- How will the data be licensed for reuse?
- If you are using third-party data, how do the permissions you have been granted affect licensing?
- Will data sharing be postponed / restricted e.g. to seek patents?

Guidance:

State who will own the copyright and IPR of any new data that you will generate. For multi-partner projects, IPR ownership may be worth covering in a consortium

agreement. If purchasing or reusing existing data sources, consider how the permissions granted to you affect licensing decisions. Outline any restrictions needed on data sharing e.g. to protect proprietary or patentable data.

See the DCC guide: [How to license research data](#).

### **The University of Sheffield: guidance on IPR Ownership and Licencing**

Please see the University of Sheffield webpage on '[Regulatory requirements](#)' for guidance on IPR and data licensing.

The UK Data Service provides guidance on '[Rights relating to research data](#)' including [licensing](#).

The Digital Curation Centre provides guidance on '[How to license research data](#)'.

### **DCC guidance on Ethical Issues**

Questions to consider:

- Have you gained consent for data preservation and sharing?
- How will sensitive data be handled to ensure it is stored and transferred securely?
- How will you protect the identity of participants? e.g. via anonymisation or using managed access procedures

Guidance:

Investigators carrying out research involving human participants must ensure that consent is obtained to share data. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. Ethical issues may affect how you store data, who can see/use it and how long it is kept. You should show that you're aware of this and have planned accordingly.

See UKDS guidance on [consent for data sharing](#)

### **The University of Sheffield: guidance on Ethical Issues**

For guidance on ethical and legal issues please see the University of Sheffield webpage on '[Regulatory requirements](#)'.

See also the UK Data Service webpages on '[Legal and ethical issues](#)'.

### **Preservation Plan**

**State the long-term preservation plan for the dataset**

#### **CRUK Guidance**

Once the funding for a project has ceased researchers should preserve all data resulting from that grant to ensure that data can be used for follow-up or new studies. Cancer Research UK expects that data be preserved and available for sharing with the science community for a minimum period of five years following the end of a

research grant.

## DCC guidance on Preservation Plan

Questions to consider:

- What is the long-term preservation plan for the dataset? e.g. deposit in a data repository
- Will additional resources be needed to prepare data for deposit or meet charges from data repositories?

Guidance:

Researchers should consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving.

If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.

## The University of Sheffield: guidance on Preservation Plan

**Note:** For guidance see the University of Sheffield webpages on '[Preserving your data](#)' and '[Data repositories](#)'.

Long term preservation and access may be best managed by using a specialist data repository. Your funder may specify a data repository to use, such as [UK Data Service ReShare](#), [NERC Data Centres](#) or [Archaeology Data Service](#). Alternatively, look in [re3data.org](#), at [Wellcome Trust - Data repositories and database resources](#) and at [BBSRC supported resources](#) to find an appropriate repository. If no suitable repository is available you may [deposit data in ORDA](#), the University of Sheffield data repository. Alternatively, if you need to regulate users' access through 'Data sharing agreements', data may be retained in the University's research storage infrastructure and [registered in ORDA](#).

**Suggested text in all cases:** *"Data will be archived in line with the University of Sheffield's Research Data Management Policy, which is a component of the University's Policy on Good R&I Practices (the 'GRIP' Policy)."*

**Where data is in paper format:** *"Data collected in paper form will be routinely digitised and the paper form disposed of / stored for at least 10 years at our universities in secured areas."*

**For data deposited in external data repositories:** *"Research data selected for long-term preservation and sharing will be deposited in [name of repository/weblink]. The [name of repository] is openly accessible and searchable and will guarantee preservation of these data for ten years or more. Metadata records describing these data will be created in ORDA, the University of Sheffield research data registry and repository"*

**Where some research data are being deposited in ORDA:** *"Data that are not*

*deposited in [name of repository/weblink] will be deposited in ORDA, a repository and registry of research data produced at the University of Sheffield, which will preserve data for ten years or more.”*

**Where data is deposited in ORDA only:** *“Data selected for long-term preservation and sharing will be deposited in ORDA, a repository and registry of research data produced at the University of Sheffield, which will guarantee preservation for ten years or more.”*

**Where data is being retained locally, but not made ‘openly’ accessible:** *“Data selected for long-term preservation and sharing will be stored on centrally provisioned University of Sheffield virtual servers and research storage infrastructure (<https://www.sheffield.ac.uk/cics/research>) for at least ten years. Records of these data will be published in ORDA, a registry of research data produced at the University of Sheffield.”*