Wellcome Trust Template

Plan Name Wellcome Trust Template

Principal Investigator / **Researcher** TUoS Researcher

Funder -

Institution University of Sheffield

Data outputs

What data outputs will your research generate?

Wellcome Trust Guidance

Any data that is shared should be of a sufficiently high quality and in a format that enables it to be used effectively.

We recognise that in some cases it may not be appropriate for researchers to share data outputs. If you don't intend to share your data, you must justify your reasons .

Data should be shared in accordance with recognised data standards, where these exist, and in a way that maximises opportunities for data linkage and interoperability. BioSharing is one directory of available data standards.

You should:

- provide sufficient metadata to enable the dataset to be discovered, interpreted and used by others
- adopt agreed best practice standards for metadata provision where these are in place.

When developing a data management and sharing plan, you should consider and briefly describe:

- the types of data the proposed research will generate
- which data will have value to other research users and could be shared
- the data formats and quality standards that will be applied to enable the data to be shared effectively.

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 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 sharing

When do you intend to share your data?

Wellcome Trust Guidance

You must state the timescale for sharing datasets of value. This should take account of any recognised standards of good practice in your research field.

We recognise that researchers have the right to a reasonable (but not unlimited) period of exclusive use for the research data they produce.

As a minimum, you should make the data underpinning research papers available to other researchers at the time of publication providing this is consistent with:

- any ethics approvals and consents that cover the data
- any valid restrictions relating to intellectual property.

We encourage researchers to make this data openly available wherever feasible via recognised subject data repositories, or general community repositories (eg Dryad, Zenodo and FigShare). Please read our requirements for publishing Wellcome-funded research papers [PDF 49KB] for more information.

We encourage researchers to increase opportunities for timely and responsible pre-publication sharing of datasets. Where appropriate, you may use publication moratoria to facilitate pre-publication data sharing with other researchers, while protecting your right to first publication.

Any such restrictions on data use should be reasonable, transparent and in line with established best practice in the respective field.

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]."*

Where will your data be made available?

Wellcome Trust Guidance

You should deposit data in recognised data repositories for particular data types where they exist, unless there's a compelling reason not to do so. Find out which repositories may be appropriate.

If you intend to create a tailored database resource or to store data locally, you should ensure that you have the resources and systems in place to curate, secure and share the data in a way that maximises its value and guards against any associated risks.

You need to consider how data held in this way can be effectively linked to and integrated with other datasets to enhance its value to users.

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. Some funders specify a data repository to use, such as the UK Data Service ReShare, NERC Data Centres or the Archaeology Data Service.

To find an appropriate repository, look in:

- Wellcome Trust - Data repositories and database resources
- re3data.org
- BBSRC supported resources

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.

How will your data be accessible to others?

Wellcome Trust Guidance

Access procedures:
Where a managed access process is required - eg where a study involves potentially identifiable data about research participants - the access mechanisms you set up should be proportionate to the risks associated with the data. They must not unduly restrict or delay access.

You must describe any managed access procedures you're proposing in your data management and sharing plan.

Depending on the study, you may want to establish a graded access procedure. For example, less sensitive data (eg anonymised and aggregate data) is made readily available, whereas applications to access to more sensitive datasets are subject to a more stringent assessment process.

Any managed access procedures should be consistent and transparent.

In cases where a Data Access Committee is required to assess applications to access data, the committee should include individuals with appropriate expertise who are independent of the project.

The Expert Advisory Group on Data Access has set out key principles for developing data access and governance mechanisms, to which applicants should refer.

Citing data outputs:
We encourage all researchers to attain digital object identifiers (DOIs) or other form of persistent identifiers for their data outputs to enable their re-use to be cited and tracked.

The DataCite initiative provides a key route through which DOIs are assigned to datasets. Many repositories assign DOIs on deposition.

Where appropriate, you may also publish a 'data paper' or other form of publication, so data users can formally cite their use of the resource.

Enabling discoverability:
Where a database resource is being developed as part of a funded activity, you should take reasonable steps to ensure that potential users are made aware of its availability. Your plan should outline your approach for enabling discoverability of your data.

DCC guidance on Discovery by Users
Questions to consider:

- How will potential users find out about your data?
- Will you provide metadata online to aid discovery and reuse?

Guidance:

Indicate how potential new users can find out about your data and identify whether they could be suitable for their research purposes. For example, you may provide basic discovery metadata online (i.e. the title, author, subjects, keywords and publisher).

The University of Sheffield: guidance on Discovery by Users

Note: All research data selected for long-term preservation should be registered in the University of Sheffield research data repository ORDA. A metadata record should be created in ORDA irrespective of whether the data files are deposited in ORDA or in another repository. Research data in non-digital formats and digital data that cannot be made accessible or requires controlled access should also be registered in ORDA. Please see the University of Sheffield webpage on 'ORDA user guidance'.

Suggested text in all cases: *"Records of datasets will be published in ORDA, the University of Sheffield's registry of research data produced at the University, which will issue DataCite DOIs for registered datasets and promote discovery."*

DCC guidance on Managed Access Procedures
Questions to consider:

- Will access be tightly controlled or restricted? e.g. by using data enclaves / secure data services
- Will a data sharing agreement be required?
- How will the data be licensed for reuse?

Guidance:

Indicate whether external users will be bound by data sharing agreements, licenses or end-user agreements. If so, set out the terms and key responsibilities to be followed. Note how access will be controlled, for example by the use of specialist services. A data enclave provides a controlled secure environment in which eligible researchers can perform analyses using restricted data resources. Where a managed access process is required, the procedure should be clearly described and transparent.

The University of Sheffield: guidance on Managed Access Procedures

Note: Restrictions on the release of data may be allowed, to protect confidentiality and for other ethical and legal considerations. Access to and use of Sensitive and confidential data can be restricted and regulated using end user licenses, data

sharing agreements or by using a data enclaves.

Whatever form of publishing is used, research data needs to be licensed to indicate what users may or may not do with the data. Data repositories will indicate what licenses are available for the data they house. More information is available from the Digital Curation Centre webpage 'How to license research data'.

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

Suggested text: *"The University of Sheffield's Good Research and Innovation Practice (GRIP) Policy follows RCUK principles for data sharing (http://www.rcuk.ac.uk/research/datapolicy/)"*

Suggested text where a Data sharing agreement is required: *"External users will be bound by data sharing agreements as specified by the [name of funder] Data Sharing Policy."*

Suggested text where an external collaborator is involved: *"Data sharing agreements will be put in place with [name of collaborator], who will be a primary re-user of data"*

Faculty of Medicine, Dentistry & Health: guidance on Managed Access Procedures

Access to and use of Sensitive and confidential data can be restricted and regulated using end user licenses or data sharing agreements.

MRC, Wellcome Trust and Cancer Research UK provide guidance on data sharing agreements in 'Good Practice Principles for Sharing Individual Participant Data from Publicly Funded Clinical Trials'.

MRC provides information on data access governance, facilitating data access and data sharing agreements in 'MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies' and 'Principles for access to, and use of, MRC funded research data'.

The University of Sheffield ScHARR Information Governance Policy provides guidance on information sharing and an example data sharing agreement.

Are any limits to data sharing required?

Wellcome Trust Guidance

For some research, delays or limits on data sharing may be necessary to safeguard research participants or to ensure you can gain intellectual property protection.

But restrictions should be minimised as far as possible and set out clearly in data management and sharing plans where these are required.

Safeguarding research participants:
For research involving human subjects, data must be managed and shared in a way that's fully consistent with the terms of the consent under which samples and data were provided by the research participants.

For prospective studies, consent procedures should include provision for data sharing

in a way that maximises the value of the data for wider research use while providing adequate safeguards for participants. Proposed procedures for data sharing should be set out clearly, and current and potential future risks explained to participants.

When designing studies you must ensure that you protect the confidentiality and security of human subjects through appropriate anonymisation procedures and managed access processes.

Systems should safeguard participants but also be proportionate to the data's level of sensitivity and associated risk. They should not unduly inhibit responsible data sharing for legitimate research uses.

Intellectual property:
As a funded researcher, you need to ensure that any intellectual property that comes from your research is suitably protected and managed, in line with our intellectual property and patenting policy.

Delays or restrictions on data sharing which may be appropriate to gain intellectual property protection or to further develop a technology for public benefit should be minimised as far as possible.

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'.

Data preservation

How will key datasets be preserved?

Wellcome Trust Guidance

You need to consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant.

If your proposal is to create a bespoke data resource or to store data locally, rather than to use a recognised data repository, your data management plan should state how you expect to preserve and share the dataset when your funding ends.

We're happy to discuss issues relating to longer-term preservation and sustainability to maximise the long-term value of key research datasets.

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. Some funders specify a data repository to use, such as the UK Data Service ReShare, NERC Data Centres or the Archaeology Data Service.

To find an appropriate repository, look in:

- Wellcome Trust - Data repositories and database resources
- re3data.org
- BBSRC supported resources

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."*

Required resources

What resources will you require to deliver your plan?

Wellcome Trust Guidance

You should carefully consider what resources you may need to deliver your plan and outline where dedicated resources are required.

Examples of resources you can request include:

People and skills:

- support for one or more dedicated data manager or data scientist (on a full- or part-time basis)
- specific data management or analysis training for research or support staff that is needed to deliver the proposed research.

We don't usually consider costs for occasional or routine support from institutional data managers or other support staff.

Data storage and computation:

- any dedicated hardware or software that is required to deliver your proposed research
- the cost of accessing a supercomputer or other shared facilities.

We would usually expect costs associated with routine data storage to be met by the institution. We will only consider storage costs associated with large or complex datasets which exceed standard institutional allowances.

Data access:

- the reasonable costs of operating a data access committee or other form of managed access mechanism over the lifetime of the award
- the costs of ingesting secondary data from users
- costs associated with accessing data from others researchers that you need to take forward your proposed research

Data deposition and preservation:

- the data ingestion costs for recognised subject repositories
- costs for deposition in unstructured repositories (eg FigShare, Dryad and Zenodo) where no recognised subject repository exists.

If no repository is suitable, we may consider ingestion costs for institutional repositories.

We don't usually consider estimated costs for data curation that extend beyond the lifetime of the grant. But we're willing to discuss how we can help support the long-term preservation of high-value data resources on a case-by-case basis.

DCC guidance on Resourcing

Questions to consider:

- What additional resources are needed to deliver your plan?
- Is additional specialist expertise (or training for existing staff) required?
- Do you have sufficient storage and equipment or do you need to cost in more?
- Will charges be applied by data repositories?
- Have you costed in time and effort to prepare the data for sharing / preservation?

Guidance:

Carefully consider any resources needed to deliver the plan. Where dedicated resources are needed, these should be outlined and justified. Outline any relevant technical expertise, support and training that is likely to be required and how it will be acquired. Provide details and justification for any hardware or software which will be purchased or additional storage and backup costs that may be charged by IT services.

Funding should be included to cover any charges applied by data repositories, for example to handle data of exceptional size or complexity. Also remember to cost in time and effort to prepare data for deposit and ensure it is adequately documented to enable reuse. If you are not depositing in a data repository, ensure you have appropriate resources and systems in place to share and preserve the data.

See UKDS guidance on costing data management.

The University of Sheffield: guidance on Resourcing

The University of Sheffield research data storage facility allocates 10TB storage free to research groups during the lifetime of a project. If a larger quota is required then this will involve charges. Long-term archiving of data may involve charges also. Get in touch with CiCS to discuss your requirements and get a quote at https://www.sheffield.ac.uk/cics/support/help.

ORDA, the University of Sheffield research data repository is free to use. You should enquire about charges made by other data repositories you intend to use.