BBSRC Template

Plan Name BBSRC Template

Principal Investigator / **Researcher** TUoS Researcher

Funder -

Institution University of Sheffield

Data areas and data types

Outline the volume, type and content of data that will be generated e.g. experimental measurements, models, records and images

BBSRC Guidance

BBSRC recognises that effective data sharing is already practiced in certain areas and expects this to continue. BBSRC supports, either directly or indirectly, a number of such resources. Data sharing in other areas is also expected where there is a strong scientific case and where it is cost effective.

BBSRC has identified a number of areas where there is a particularly strong scientific case for data sharing. These are:

- Data arising from high volume experimentation
- Low throughput data arising from long time series or cumulative approaches
- Models generated using systems approaches

BBSRC expects data sharing to take place in these areas.

DCC guidance on Data Description

Questions to consider:

- What data will you create?

Guidance:

Give a brief description of the data that will be created, noting its content and coverage

The University of Sheffield: guidance on Data Description

Please see the University of Sheffield webpage 'What is research data?' for guidance.

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.

Standards and metadata

Outline the standards and methodologies that will be adopted for data collection and management, and why these have been selected

BBSRC Guidance

Standards are fundamental to effective data sharing. These can include standards for administrative processes, as well as for methodologies relating to data management and data formats. Researchers are expected to make use of current guidance and information on best practice.

It is expected that, in order to maximise the potential for re-use of data, BBSRC researchers should generate and manage data using existing widely accepted formats and methodologies where available. Data released for sharing should be validated and verified in line with accepted best practice and be of high quality. Data should be accompanied by the contextual information or documentation (metadata) needed to provide a secondary user with any necessary details on the origin or manipulation of the data in order to prevent any misuse, misinterpretation or confusion. Where standards for metadata exist, it is expected that these should be adhered to.

BBSRC encourages community development of standards where these do not currently exist or are not widely accepted and provides funding mechanisms for

support of this type of activity.

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.

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 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.

Relationship to other data

State the relationship to other data available in public repositories

DCC guidance on Relationship to Existing Data

Questions to consider:

- What is the relationship to existing data e.g. in public repositories?
- How does your data complement and integrate with existing data?

Guidance:

Consider the relationship between the data that you will capture and existing data available in public repositories or elsewhere.

Secondary Use

Outline the further intended and/or foreseeable research uses for the completed dataset(s)

BBSRC Guidance

BBSRC supports the view that those enabling sharing should receive full and appropriate recognition by funders, their academic institutions and new users for promoting secondary research.

Where data are shared through a third party resource or databases, secondary users should acknowledge the source of data. Where data are shared directly from the originator, depending on the level of usage and collaboration either joint authorship or acknowledgement to the data originator may be appropriate. It is also important to ensure that researchers and their research institutions are protected against claims that application of their data led to wrong conclusions/decisions by others: any use made of any data generated by third parties would not come with a warranty of its quality.

Furthermore, BBSRC expects that researchers accessing data have responsibilities to preserve data confidentiality and to observe the ethical and legal obligations pertaining to the data.

DCC guidance on Expected Reuse

Questions to consider:

- Who may be interested in using your data?
- What are the further intended or foreseeable research uses for the data?

Guidance:

You should think about the possibilities for reuse of your data in other contexts and by other users, and connect this as appropriate with your plans for dissemination and Pathways to Impact. Where there is potential for reuse, you should use standards and formats that facilitate this. Where possible outline the types of users you expect and estimate numbers.

The University of Sheffield: guidance on Expected Reuse

The potential of research data for reuse must be considered when planning long-term curation. The Digital Curation Centre provides useful advice about data selection and appraisal. The NERC data value checklist provides guidance on determining long-term value.

Methods for data sharing

Outline the planned mechanisms for making these data available, e.g. through deposition in existing public databases or on request, including access mechanisms where appropriate

BBSRC Guidance

BBSRC recognises that different approaches to data sharing will be required in different situations and considers that it is most appropriate for researchers to determine their own strategies for data sharing and outline these within their research grant proposal(s). Applicants should consider where, how, and to whom their data should be made available.

In addition, data sharing practices will change as areas of research develop and become more mature. This can be observed by looking at the areas of sequencing (i.e. well established mechanisms in place), microarrays (i.e. standards developed and being implemented) and systems biology (i.e. databases currently not well developed). Consideration should be given to what constitutes good practice in emerging areas of research.

It is expected that data sharing strategies will fall into the two broad categories below.

Data Sharing via a 3rd Party

Data sharing via deposition in an existing database, repository or other community resource is expected where possible and researchers are encouraged to share data through mechanisms affording the widest availability for generating added value and enabling re-use.

Researchers are encouraged to use existing infrastructure to facilitate data sharing where possible. BBSRC funds or otherwise supports a number of such resources. Where no such resources exist, applicants may consider sharing data via other third party mechanisms such as journal websites and / or open access repositories, many of which are now able to capture and share data underpinning publications.

Direct Data Sharing: from Originator to Others

This method of data sharing may be appropriate for areas where suitable third party mechanisms are not available. Researchers are expected to ensure that data are maintained for a period of 10 years after the completion of the research project in suitable accessible formats using established standards where possible such that the data can be made available on request in line with BBSRC guidance on good scientific practice. This may lead to collaboration between the new user and the original data creators, with the responsibilities and rights of all parties agreed at the outset.

Other mechanisms for data sharing may be used where appropriate. These could include sharing data within closed communities or a combination of methods for different datasets. Specific access mechanisms could be appropriate for example where there are ethical considerations, a need to protect confidential data, or other reasons for limiting access.

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. Some funders specify a data repository to use, such as UK Data Service ReShare, NERC Data Centres or Archaeology Data Service. To find an appropriate repository, look in

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

Proprietary data

Outline any restrictions on data sharing due to the need to protect proprietary or patentable data

BBSRC Guidance

In instances where BBSRC and a commercial partner jointly fund academic research work (for example LINK projects) there may be some restrictions over releasing data. Any such restrictions on data sharing due to co-funding arrangements should be set out in the “statement on data sharing” section of an application and will be considered when a grant application is peer reviewed. Applicants should also ensure they have obtained necessary clearances from relevant collaborators with regards to the content of the proposal including the data sharing plan in line with the BBSRC Research Grants Guide.

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.

DCC guidance: 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'.

Timeframes

State the timescales for public release of data

BBSRC Guidance

The value of data often depends on timeliness. Researchers have a legitimate interest in benefiting from their own time and effort in producing data, but not in prolonged exclusive use of these data. BBSRC expects that all data (with accompanying metadata) should be shared in a timely fashion as soon as it is verified. It is expected that timely release would generally be no later than the release through publication of the main findings and should be in-line with established best practice in the field. Where best practices does not exist release within three years of generation of the dataset is suggested as a guide.

The timescale for release for the data may differ for several reasons, depending on the nature of the data. These reasons may include:

- **Scientific Area:** Researchers are expected to make data available in-line with established practices within the relevant research community. Examples include:
    - Crystallography (Protein Data Bank) - the community has agreed a maximum 12-month delay between publishing the first paper on a structure and making coordinates public for secondary use.
    - Sequencing (EMBL Nucleotide Sequence database) – submitted data can be withheld from public access until publication of results but no later.
    - Metabolomics (MeT-RO) – Up to a six-month delay in publication can be requested.
    - Arabidopsis microarray data (NASC Affymetrix service) – all data are made available after a maximum one-year confidential period.
- **Intellectual Property (IP) issues and potential for commercialisation of research outputs:** New knowledge generates patentable ideas. BBSRC is also driving a policy of Knowledge Transfer and strongly encourages the commercialisation of IP through various initiatives. BBSRC recognises the need for periods of exclusive use of data but considers that commercialisation of research does not preclude data sharing and should not unduly delay or prevent data sharing. Any IP issues or plans for commercialisation should be highlighted in the case for support of the grant application.
- **Length or scope of research project:** Data from large studies may be released in waves as they become available or as they are published.

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]."*

Formats

State the format of the final dataset

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.