



BBSRC DATA SHARING POLICY

Version 1.1

June 2010



This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Contents

Background to BBSRC's Data Sharing Policy	3
Background and Context	3
<i>General</i>	3
<i>Scientific</i>	3
BBSRC's Position	3
BBSRC Data Sharing Policy Statement	5
Data Sharing Areas	5
Policy Implementation 1: Integrating Data Sharing into Existing Support and Monitoring Processes	7
Data Sharing Statements.....	7
<i>Research Grant Proposals</i>	7
<i>Assessment of the Data Sharing Statement</i>	7
Data Sharing Plans.....	7
<i>Areas for data sharing and types of data</i>	8
<i>Standards and Metadata</i>	9
<i>Methods of Data Sharing</i>	10
<i>Timeframe for Data Sharing</i>	11
<i>Secondary Use of Data</i>	11
<i>Proprietary Data</i>	12
<i>Data Management: the Project Life Cycle</i>	12
<i>Funding for Sharing of Research Data</i>	12
<i>Monitoring of Data Sharing through Final Reports</i>	13
<i>Ethical Considerations</i>	13
BBSRC-sponsored Institutes	13
Policy Implementation 2: Mechanisms to Facilitate Data Sharing	14
Funding Opportunities	14
Other Related Activities	15
Other Research Councils Data Sharing Policies	15
UK PubMed Central.....	15

Background to BBSRC's Data Sharing Policy

Background and Context

General

The principles of data sharing are widely recognised and underpin many international activities. A report by the Organisation for Economic Co-operation and Development (OECD) on "[Promoting Access to Public Research Data for Scientific, Economic and Social Development](#)" highlights the following principles:

- Publicly-funded research data are a public good, produced in the public interest
- Publicly-funded research data should be openly available to the maximum extent possible

The report concludes that widespread data sharing will enable researchers, empower citizens and convey tremendous scientific, economic, and social benefits. The BBSRC supports this view.

Scientific

BBSRC sponsors a wide range of scientific research that generates large volumes of data. This includes, for example, information about protein structure, DNA sequencing and proteomics as well as data arising from imaging, agricultural, environmental research and species information.

Historically, BBSRC has provided funds to establish 'high throughput' approaches to experimentation (e.g. the Investigating Gene Function Centres), and more recently we have recognised the importance of sustainability of such resources (e.g. through the Bioinformatics and Biological Resources Fund). New research methods and approaches mean that biological scientists are generating increasingly large volumes of research data. Such data are important not only for the researchers originating the work but also to the wider scientific community, which may wish to examine and use these datasets to underpin other investigations. Re-use of data can lead to new scientific understanding and examples of this already exist, e.g. the rapid re-analysis of genomic sequences to reveal the [origins and evolution of the H1N1 virus](#).

BBSRC's Position

BBSRC recognises the importance of contributing to the growing international efforts in data sharing. BBSRC is committed to getting the best value for the funds we invest and believes that making research data more readily available will reinforce open scientific enquiry and stimulate new investigations and analyses. BBSRC supports the view that data sharing should be led by the scientific community and driven by scientific need. It should also be cost effective and the data shared should be of the highest quality. Members of the community are expected and encouraged to practise and promote data sharing, determine standards and best practice, and create a scientific culture in which data sharing is embedded. BBSRC will provide support and funding to facilitate this.

Ownership of the data generated from the research that BBSRC funds resides with the investigators and their institutions. Therefore, the data sharing policy will be implemented by:

- Integration into the processes of supporting and monitoring research (including both research grants and the core strategic grants provided to the BBSRC-sponsored institutes)
- Providing mechanisms to facilitate and encourage data sharing in the bioscience community

BBSRC Data Sharing Policy Statement

BBSRC expects research data generated as a result of BBSRC support to be made available with as few restrictions as possible in a timely and responsible manner to the scientific community for subsequent research. Applicants should make use of existing standards for data collection and management and make data available through existing community resources or databases where possible. In line with the [BBSRC Statement on Safeguarding Good Scientific Practice](#), data should also be retained for a period of ten years after completion of a research project.

BBSRC recognises that different fields of study will require different approaches. What is sensible in one scientific or technological area may not work in others; therefore the policy aims to achieve the sharing of data in an appropriate manner and not to be overly prescriptive. Researchers are required to adhere to any relevant regulatory requirements, including those relating to the ethical use of data.

BBSRC recognises the importance of data quality and provenance. Data should, wherever appropriate and possible, be accompanied by contextual information or documentation (metadata) 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.

The value of data often depends on timeliness. Researchers have a legitimate interest in benefiting from their own time and effort in producing the data but not in prolonged exclusive use of these data. Timescales for data sharing will be influenced by the nature of the data but **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 field¹. BBSRC considers data sharing to be an important activity and whilst recognising the need to safeguard Intellectual Property and to protect opportunities for commercialisation of research outputs considers that this should not unduly delay or prevent data sharing.

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.

Data Sharing Areas

BBSRC recognises that effective data sharing is already practised in certain areas and expects this to continue. Through consultation, BBSRC has identified 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. Data sharing in other areas is also encouraged where there is strong scientific need and where it is cost effective. At

¹ Where best practice does not exist, release of data within three years of its generation is suggested.

regular intervals, BBSRC will review scientific and technological developments in the biosciences to update the areas highlighted.

BBSRC reserves the right to implement a more prescriptive approach to data sharing for research initiatives (particularly those involving large scale collaborative approaches) or where BBSRC is supporting a community resource, including software. Such an approach has been adopted by the US NIH for its Glue Grants program (for more details of the NIH data sharing policy see http://grants.nih.gov/grants/policy/data_sharing and the [Glue Grants programme](#) and by other UK Research Councils (for example for the [Cross-Council Rural Economy and Land Use Programme](#)).

Policy Implementation 1: Integrating Data Sharing into Existing Support and Monitoring Processes

Data Sharing Statements

Research Grant Proposals

All applications seeking research grant funding from BBSRC must submit a statement on data sharing. This should include concise plans for data management and sharing as part of research grant proposal **or** provide explicit reasons why data sharing is not possible or appropriate. The “statement on data sharing” will be included as an additional page in the case for support. Applicants **must not** utilise this space allocation for any other purpose - as this will result in rejection of the proposal.

Assessment of the Data Sharing Statement

An application’s “Data Sharing Statement” will be assessed by reviewers and BBSRC Responsive Mode Committees or Assessment Panels. The statement will be considered separately from the scientific excellence of the proposed research; however, an application’s credibility will suffer if peer review agrees the statement is inappropriate. In the case where a highly rated proposal has an inappropriate data sharing statement Committees and Panels may choose to offer conditional awards and/or provide specific feedback to the applicants. Appropriate plans are expected to be those where the proposed data sharing activities are in-line with current best practice in the field and both the scientific and cost benefits are considered.

Guidance on assessment of the data sharing statements will be included in the standard guidance provided to reviewers and Committee members.

Data Sharing Plans

BBSRC recognises that data sharing plans will vary according to the type of data collected. Data sharing should be driven by scientific benefit and should also be cost effective. Data should be shared using established standards and existing resources where this is possible. Applicants may wish to include details of:

- *Data areas and data types* - the volume, type and content of data that will be generated e.g. experimental measurements, models, records and images;
- *Standards and metadata* - the standards and methodologies that will be adopted for data collection and management, and why these have been selected;
- *Relationship to other data available in public repositories*;
- *Secondary use* - further intended and/or foreseeable research uses for the completed dataset(s);
- *Methods for data sharing* - planned mechanisms for making these data available, e.g. through deposition in existing public databases or on request, including access mechanisms where appropriate;
- *Proprietary data* - any restrictions on data sharing due to the need to protect proprietary or patentable data;
- *Timeframes* - timescales for public release of data;
- *Format of the final dataset*.

Further guidance on these points is provided in the sections that follow.

Areas for data sharing and types of data

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. Assessment Committees and Panels will give particular consideration to the 'Data Sharing Statement' when reviewing research grant proposals in these areas. At regular intervals, BBSRC will review scientific and technological developments in the biosciences to update the areas highlighted and to ensure data sharing is of benefit to the scientific community. In 2009, BBSRC decided that software arising from funded grants should sit within the framework of the data sharing policy.

There may be areas where there is unlikely to be value in sharing data, where there are limited scientific benefits, experiments are readily repeatable, and/or the costs of making data available for sharing are high. These could include, for example, commonly collected low throughput data such as gene disruption and enzyme kinetics. In such cases, applicants are encouraged to explore the most appropriate methods by which their data could be shared.

Data from High Volume Experiments

For these purposes, "high volume" experiments are defined as those generating a dataset or datasets consisting of 100s of measurements generated in parallel or near-parallel from a single experimental sample that can be captured and stored in a readily accessible electronic format. Technologies currently expected to generate high volume datasets could include "omics" technologies, sequencing etc.

Illustrative examples would be:

- Drosophila genome database, Flybase (<http://flybase.bio.indiana.edu/>)
- Nucleotide sequence database EMBLbank (<http://www.ebi.ac.uk/embl/>)
- Arabidopsis information resources (<http://arabidopsis.info/>)

Low Throughput Data from Long Time Series or Cumulative Approaches

Long Term Studies (LTS) are used to research the changes in a complex biological system responding over time to environmental or other factors whose influences are very difficult to predict, or at such large temporal or spatial scales, that they only emerge through sustained monitoring. The duration of an LTS would be expected to significantly longer than the duration of a single responsive-mode grant, typically over time scales in excess of 10 years, and over multiple seasons and generation times for the organisms under study.

LTS typically make many standardised measurements and collect samples at regular intervals, creating both data and biological sample resources that can be subjected to retrospective analysis.

The unique nature of an LTS, in terms of its place in time and environment, means that the data cannot be made up for, replaced, or duplicated. Furthermore, while LTS studies are running, the data set and its analysis cannot be considered complete because the addition of new data can lead to different findings.

Because of these unique features of LTS, it is recognised that active stewardship is required to sustain their scientific integrity and long-term continuity. It is also recognised that added value of such resources is often gained from their utilisation for purposes other than those for which they were originally designed

Illustrative examples would be:

- The Rothamsted Classical Experiments (e.g. Broadbalk and Park Grass) (www.rothamsted.bbsrc.ac.uk/resources/ClassicalExperiments.html)
- The Rothamsted Insect Survey (www.rothamsted.bbsrc.ac.uk/insect-survey/)
- The UK Butterfly Monitoring Scheme (<http://www.ukbms.org/>)
- The UK Environmental Change Network (www.ecn.ac.uk/)
- US Long-Term Ecological Research sites (www.lternet.edu/)

Note: BBSRC expects that any biological resources accompanying the data should also be made available through deposition at the most appropriate resource (for example National Arabidopsis Stock Centre; The National Collection of Yeast Cultures at Institute of Food Research). If no collection (either UK or International) will accept the organisms it is expected that the grant holders themselves should make them available to the scientific community for a minimum period of 2 years following publication of any paper describing those organisms. See the [BBSRC Research Grant Guide](#).

Models Generated using Systems Approaches

Models, generated and iterated using integrative systems approaches to biology, are in themselves a valuable and shareable output of research, as important as the data they generate. These models are essential in underpinning the conclusions of systems biology experiments and, as such, should be made freely available to other researchers wishing to reproduce the experiments.

Wherever possible, such models should be submitted to an appropriate repository, where they are made available to the broader scientific community. An example of such a repository is the BioModels Database (www.ebi.ac.uk/biomodels-main), hosted at the EBI which will also provide illustrative examples of the models discussed.

Standards and Metadata

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.

Methods of Data Sharing

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.

Timeframe for Data Sharing

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.

Secondary Use of Data

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.

Proprietary Data

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](#).

Data Management: the Project Life Cycle

A dataset typically has a longer lifespan than the research project which creates it. The lifespan of a research project can be formally defined by the period for which it is funded. Researchers may continue to use datasets after initial funding has ceased, and follow-up projects may subsequently be funded to continue to analyse or add to the datasets. In addition, new research projects may re-use data generated by other researchers, particularly for comparative studies or analyses. Data generated through a research project may be shared, or archived, or both. Not all data that are appropriate for sharing should be archived, and *vice versa*.

To enable data sharing, it is expected that appropriate data management strategies should be in place throughout the research project. Activities involved in preparing a dataset for sharing (particularly *via* submission to a public resource) should ideally be done within the funding period of the project. If not, knowledge, staff and motivation may be lost and the work may not be carried out to adequate standards.

Longer term data storage, or archiving, is outside the scope of this policy. Researchers should note that the BBSRC “[Safeguarding Good Scientific Practice](#)” document states that it expects primary data to be securely held for a period of ten years after completion of a research project, and institutions receiving funding from BBSRC to have guidelines setting out responsibilities and procedures for keeping data. Researchers should therefore ensure they retain a local copy of any data submitted to third party resources. This should be maintained according to institutional procedures.

Funding for Sharing of Research Data

BBSRC recognises that data sharing has time and cost implications. Funding to support the management and sharing of research data (for example staffing, physical resources such as storage and networking capability) can be requested as part of the full economic cost of a research project. Further details on full economic costing can be found in the [BBSRC Research Grants Guide](#).

Monitoring of Data Sharing through Final Reports

Adherence to the proposed data management and sharing strategies set out in a funded proposal will be monitored through the Final Report assessment procedure. Consideration of the data sharing activities will be built into the Final Report score provided to Assessment Committees. This information, as part of an applicant's track record, may be taken into account when assessing future proposals.

Ethical Considerations

Researchers are required to adhere to any relevant regulatory requirements including those relating to the ethical use of data.

BBSRC-sponsored Institutes

BBSRC supports a large amount of research activity through provision of funding *via* institute strategic programme grants to the BBSRC-sponsored research institutes. Researchers at these Institutes will be expected to comply with the principles set out in the policy statement. Monitoring of data sharing practices in the Institutes will be carried out through the Institute Assessment Exercise. Further guidance on implementation within the BBSRC-sponsored institutes will be provided separately.

Policy Implementation 2: Mechanisms to Facilitate Data Sharing

BBSRC will facilitate and encourage data sharing by providing:

- Funds (through the schemes detailed below) to support
 - development of standards and software tools which enable data sharing;
 - community resources and facilitate development of data sharing approaches in specific communities;
 - data sharing activities as part of the full economic cost of research projects.
- Information and guidance to applicants including information about existing standards, guidelines, databases and resources that may be relevant.
- Support for relevant training activities.

Funding Opportunities

The following funding schemes are available to provide support for data sharing activities:

- **Responsive mode research grants:** BBSRC recognises that data sharing has time and cost implications. Funding to support the management and sharing of research data (for example staffing, physical resources such as storage and networking capability) can be requested as part of the full economic cost of a research project. Further details on full economic costing can be found in the [BBSRC Research Grants Guide](#).
- [Special Opportunities](#),
 - *Tools and Resources Development Fund:* This is a pump-priming fund intended to support new and innovative projects, but also with provision to fund projects that promote data sharing or data standards
 - *Bioinformatics and Biological Resources Fund:* A fund designed to support the establishment, maintenance and enhancement of community resources required by bioscientists
- [Strategic Priority in Technology Development](#).

Other Related Activities

Other Research Councils Data Sharing Policies

A comparison of the data sharing statements issued by other major UK science funders has been performed by the Digital Curation Centre (DCC) and can be found [here](#).

UK PubMed Central

As a part of Open Access, the Wellcome Trust (www.wellcome.ac.uk/doc%5Fwtd002766.html) has with the help of the National Centre for Biotechnology Information (NCBI) at the NIH, established a manuscript submission system, through which papers accepted for publication in a peer-reviewed journal can be deposited and made accessible through the NIH life science archive, PubMed Central (PMC). BBSRC is one of the project partners, along with a number of other UK life sciences funders including: MRC, Arthritis Research Campaign, Cancer Research UK, British Heart Foundation, and JISC. Further details: <http://ukpmc.ac.uk/>.