

University of Bristol

Research Data Storage Facility (the Facility) Policy Procedures and FAQs

This FAQs should be read in conjunction with the RDSF usage FAQs -

<https://www.acrc.bris.ac.uk/acrc/RDSF-faqs.html>

When should I use the Facility?

To store research data securely - <https://www.acrc.bris.ac.uk/acrc/storage.htm>

For storage of active research data and long term storage of data from finished research projects (identify how long your funder requires the data to be kept.) Section *What should I use the Facility for?* provides more detail.

The RDSF is sited across 3 machine rooms and data can be mirrored across separate geographical locations. Data is stored on disk with daily backup to tape; a tape archive is planned for 2015.

A PI (Data Steward) can register a Standard Project and nominate users from within the University of Bristol to use his/her storage allocation.

To share data with external researchers - <http://data.bris.ac.uk/sharing/>

The Data Steward can register a separate Collaboration Project and nominate External Data Users from other universities to join the project.

A secure protocol is used to transfer the data.

The data is 'read only' by default; it can be made 'read-write'.

Access by External Data Users will be reviewed after 3 years and can be extended as appropriate.

To publish data - <http://data.bris.ac.uk/data/>

Data can be published and assigned a Digital Object Identifier (DOI) in the University's research data repository.

The Data Steward has a separate folder within the Standard Project file space. Data can be moved into this folder as a permanent 'read-only' copy.

A minimal set of metadata is published with the data.

It is planned to link the research data repository and Pure, the University's Research Information System - <http://www.bristol.ac.uk/red/research-policy/pure/>.

To manage sensitive data - <http://www.bristol.ac.uk/infosec/uobdata/research/>.

If your research involves human participants, you will need to complete an ethical review - <http://www.bristol.ac.uk/red/research-governance/ethics/>.

Consider whether you wish to seek a FOI exemption -

<http://www.jisc.ac.uk/publications/programmerelated/2010/foiresearchdata.aspx>.

Written approval from Secretary's Office is required if unanonymised sensitive personal data is stored in the RDSF.

It is recommended that the Data Steward only registers a limited number of users to the project.

What should I use the Facility for?

The Facility is a bulk data store for the long term storage of *digital research data* and is NOT geared for high performance data processing by user programs. Users are expected to do data processing locally. If you are storing a large number of small files, which will not be accessed frequently, you should think about zipping or software archiving files to make retrieval times shorter. The RDSF is not intended for the storage of a researcher's personal data.

What is research data?

Research data is data generated by funded or unfunded research in the course of employment at the University of Bristol. It does not include data generated in the course of personal activities, desktop or mailbox backups, or data produced by non-research activities such as University administration and teaching.

Who should apply?

If the project is a funded project, normally the Principal Investigator (PI) listed in the project proposal should complete the application form. Only employees of the University of Bristol can apply. The applicant will be known as the Data Steward. All nominated users of the data by the Data Steward will be known as Data Users or in the case of a non University of Bristol employee, as External Data Users.

If the project is unfunded, then the project leader should complete the application form and become the Data Steward. The Data Steward must be employed by the University of Bristol.

Where personal data (as defined by the Data Protection Act (DPA) 1998) is generated by research in the course of employment at the University of Bristol, the University will be the Data Controller under the Act for that Data.

However, in all circumstances, the Data Steward will be held personally responsible by the University for the proper stewardship of the data stored in the Facility, including the provision of such information/metadata as is required by the Research Data Storage and Management Board to meet the University's legal and ethical obligations. The responsibilities of a Data Steward are listed in the Policy for the use of the Research Data Storage Facility sections 1.9-1.13.

How do I apply?

To apply to become a Data Steward please follow this link:

<http://www.bris.ac.uk/acrc/storage-area/apply.cgi>

A Data Steward can register two types of project: a Standard Project and a Collaboration Project. A Standard Project enables use of the RDSF by the Data Steward and other members of the University of Bristol, whom he/she authorises. A Collaboration Project is specifically for sharing data with researchers from outside the University of Bristol, whom the Data Steward authorises to have access to the RDSF.

To create a Standard Project please follow this link:

<http://www.bris.ac.uk/acrc/storage-area/apply-project.cgi>

One application should be made per project. You will be asked for details about the project and the data being created, including your confirmation that it is research data. You will need to indicate how many copies of the data you wish to hold on disk and on tape and for how long, together with confirmation of the access level of the data - <http://www.bris.ac.uk/infosec/uobdata/classifications>. You will then be given an actual or notional cost for using the Facility (see section *What are the costs of using Facility?*).

In the event of your leaving the University, your line manager will automatically become Data Steward for the project.

Where you have submitted an ethical review application to your Department or Faculty Ethics Committee concerning the research that will generate the data, a copy of that application and the outcome of the review process should be provided.

Your application will be reviewed by the Research Data Storage and Management Board and you will receive a response within ten working days.

In the event of a dispute over a decision made by the Research Data Storage and Management Board, the dispute will be escalated to the HPC Executive and then if necessary to the HPC Board, who will act as the final arbiter.

What is a Data Management Plan and why do I need to complete one?

If you have already prepared a Data Management Plan (DMP) or a similar set of guidelines, it is helpful if you submit it with your application. If you are applying for research council funding, you may well have to include a DMP in your application. The Digital Curation Centre's DMP Online tool is now available -

<http://www.dcc.ac.uk/resources/data-management-plans>.

The DMP asks a series of questions about the data you will create and what you wish to do with it and provides you with a structure to look after data in the long term.

What are the costs of using the Facility?

Whilst there are charges for using the Facility, 5TB of disk storage per Data Steward will initially be provided free of charge. This policy will be reviewed by the HPC Board annually. If you anticipate needing more than 5TB of disk storage, a request for funding should be included in your grant applications.

Details of charges for more than 5TB of storage or for tape only storage are detailed in a separate document, Costs of using the Research Data Storage Facility.

If I am involved in a joint project with another university, can I share data held in the Research Data Storage Facility with researchers from that university?

The Data Steward can set up a Collaboration Project - <http://data.bris.ac.uk/sharing/collaboration/> and then arrange for an External Data User to have access to the Collaboration Project. Access is for a maximum period of 3 years. If longer access is required, the External Data User can re-register, in agreement with the Data Steward, for a further 12 months.

This is still a pilot service and the guidance is in the course of being updated.

The default access for an External Data User is 'read only'. A Data Steward can choose to make this access 'read-write'.

Why should I use the Facility?

Home grown solutions for data storage carry significant risks; a recent study by Google on disk failure rates http://research.google.com/archive/disk_failures.pdf implied that, after two years of use, the annual disk failure rate is around 8%. This means that if data is held only on a single drive (such as an external USB drive) there is approximately a one in ten chance of data loss in the second and third years. In addition, specialist retrieval of data from damaged hardware is expensive.

The Facility provides a long term storage facility, architected using best practice, with leading edge hardware and software to maximise availability, usability, security and longevity of the data.

How long should I keep data for?

A summary of research council policy is provided by the Digital Curation Centre - <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>. For more detailed information, you should consult the relevant research council or funder.

How long can I keep data for?

The digital assets held in the Facility will be reviewed by the Research Data Storage and Management Board annually. In special cases the Research Data Storage and Management Board may recommend that some data holdings are no longer considered to have sufficient value to merit storing in the Facility. The Data Steward will be notified of any such recommendation to enable them to consider whether they wish to store the data holding elsewhere.

How many copies of my data do I need to keep?

You need to consider how easily the data could be replaced should it be lost. Is there another copy? What can go wrong? A second copy of data held in separate locations will mitigate against unlikely, but catastrophic, events such as fire or other disasters.

I have personal data – does this affect how I store it?

It is expected that any personal data (as defined by the DPA 1998) generated in the course of research will be anonymised prior to its deposit in the Facility. If a Data Steward wishes to store personal data without anonymising it, then an explanation needs to be provided for the Research Data Storage and Management Board to consider. This should normally be provided as part of your ethical review application to your Department or Faculty Ethics Committee. Where no such ethical application has been made, written clearance from the University Secretary's Office will be required. Proposed storage of unanonymised sensitive personal data (as defined by the DPA 1998) will always require written clearance from the University Secretary's Office.

Please see section 2 of the Research Data Storage Facility Terms of Use document for more detailed information. Further guidance on data protection and data security can be obtained from the Information Security web pages - <http://www.bris.ac.uk/infosec/>. Encryption can be used where access needs to be strictly controlled. If you wish to encrypt your data, we suggest that you talk to ACRC staff before completing the application form.

I have commercially confidential information – does this affect how I store it?

If a Data Steward wishes to store research data in the Facility that is subject to a confidentiality agreement, then the nature of the confidential agreement must be provided to the Research Data Storage and Management Board. Under the Freedom of Information Act 2000 (FOIA 2000), third parties may request access to information held by Public Authorities, subject to certain exemptions. Exemptions include that the information requested has been provided in confidence, or its disclosure would prejudice commercial interests. Such exemptions are interpreted strictly.

If you think an exemption to third party access under the FOIA 2000 should apply to research data that you wish to store in the Facility, then you must provide details to the Research Data Storage and Management Board of which exemption you think applies. Further information about the FOIA 2000 can be obtained from the University Secretary's Office. In addition, JISC has provided guidance written specifically for researchers on Freedom of Information and managing research records - <http://www.jisc.ac.uk/publications/programmerelated/2010/foiresearchdata.aspx>.

If you wish to encrypt your data, there is guidance available here <http://www.bris.ac.uk/infosec/uobdata/encrypt/file/>.

How do I access the data?

The storage will be exported over the university network. The two main export methods will be via CIFS (for access by Windows machines) and NFS (for access by Unix/Linux machines). It is expected that these machines in the main will be departmental servers.

How do I monitor the amount of data I am holding and whether I am close to my allocation?

The Data Steward is responsible for monitoring how the allocation is divided between his/her research projects. If using NFS or CIFS, the Data Steward can see his/her usage figures using normal tools.

What should I do with the data if I plan to be away from the University for a prolonged period of time?

If the Data Steward is absent for a prolonged period of time (e.g. on sabbatical), his/her line manager will assume the role of Data Steward temporarily and the data can be retained in the Facility. It is the responsibility of the Data Steward to notify their line manager and the ACRC staff of this change.

What happens to the data if I leave the University?

If a Data Steward leaves the University, the data will be retained in the Facility as the data belongs to the University – refer to <http://www.bris.ac.uk/research/knowtransfer/ip/ipownership.html>, regarding intellectual property (IP). Responsibility for the Data Steward's data assets will normally be transferred to his/her line manager in the first instance. All use of data must conform to the legal obligations and policies of the University. It is proposed that data will normally be kept for 12 months subject to the requirements of the Facility when someone leaves, and it will then be up to the Data Steward's line manager to decide if a request should be made to the Research Data Storage and Management Board to keep data beyond the next review.

Should I cite use of the Facility in publications?

Users are encouraged to mention use of the Facility in publications, where the data underpinning the research is stored in the Facility. The suggested template wording is *"The work was made possible in part by using the Research Data Storage Facility of the University of Bristol - <http://www.bris.ac.uk/acrc/storage>".*

How do I make data held in the Facility publicly available?

If you wish to make a subset of your data publicly available, for example to link to a publication, you can upload it from the Facility into the *data.bris* repository. The subset of data should be moved into a separate folder within the project storage allocation and then validated by the Data Steward, who can then authorise the upload to *data.bris*. Once the data is moved into the separate folder, it is 'read only' and cannot be amended.