

Restrictions to the openness of data

1 Introduction

There could be ethical, legal, contractual and commercial reasons for limiting openness to data through applying an embargo period or controlling access. Research with people and research of a commercially sensitive nature are the most likely candidates for such restrictions.

1.1 Open if possible, restricted if necessary

In general, an equilibrium should be sought that meets any legitimate reasons for constraints to openness, whilst still providing access to research data as openly as possible. For example, data emanating from research with people can usually be shared ethically and legally by:

- discussing data archiving and sharing with research participants to gain their consent for data sharing
- anonymising data where needed
- considering controlling access to data

1.2 When to assess restrictions to openness

Restrictions to the openness of data can be assessed at two points in the research lifecycle:

- at the Data Management Plan stage, e.g. as part of the grant application process and/or prior to the project commencing — this point of assessment may coincide with an ethical review of the proposed research project by the appropriate ethics committee
- at a later stage, for example when research data is being offered for deposit to the SHU Research Data Archive (SHURDA)

The advantage of assessment at the Data Management Plan stage is that decisions taken at the start of a research project may directly influence whether restrictions to openness are necessary.

First, this relates to research with people: the type of informed consent that is being sought from participants determines whether data can be preserved, shared or whether it needs to be destroyed. It is therefore essential that any decisions on data sharing are part of the ethical review by the appropriate ethics committee(s) at the start of a project in a manner that reconciles data sharing with data protection. Research Ethics Committees in the University need to be aware and have the competence to assess initial decisions on data sharing and its requirements.

Second, it is important to negotiate at the start of a project with any commercial partners about the sharing of data that (at least) underpins any publications.

In both cases, it will often be difficult if not impossible to share data if any concerns are addressed too late.

It is also important to note that external funders may require that applicants make decisions on restrictions to openness at the Data Management Plan stage. For example, RCUK has indicated that any restrictions to the openness of data should be addressed in a Data Management Plan prior to the project commencing, whether such a plan is a compulsory part of a grant application or not.¹

The advantage of assessment at a later stage, for example at the point of deposit in SHURDA, is that during the course of a research project, the need may arise for considering restrictions on openness that were not foreseen at the Data Management Plan stage.

2 Types of access restrictions available in SHURDA

- **open access** to data (after registration)
- **under embargo**, during a limited period of time, access will remain completely restricted, but metadata are available. An embargo period would give a researcher time to continue analysing and publishing from a larger dataset, even if — for example — parts of the dataset have already been opened to the public on publication of a paper. Embargoes can also be used to ensure that archived data are not published until the articles based upon them are accepted for publication or published. Embargoes can also be used to delay access whilst patents are filed or while research is commercialised
- **access upon request**, which might be required for some types of confidential or sensitive data — decisions are made by SHURDA staff on the basis of criteria formulated by the depositor and registered in SHURDA. These conditions are also displayed on the dataset's public landing page
- **non-disclosure agreements** can be used to share confidential or sensitive data with specific individuals for specific purposes and under specific terms — contact RIO if a non-disclosure agreement is required

3 Reasons for controlling access

3.1 Ethics and data protection

- The data can be classified as personal data (through which respondents can be identified) and sharing data would breach the Data Protection Act because:
 - participants did not give informed **consent** to data archiving and sharing
 - **anonymising** data is not sufficiently possible (when dealing with personal data, direct and indirect identifiers should be removed so that variables cannot be combined to reveal an individual's identity, or an organisation or business)
- Other ethical considerations: when inappropriate release of data might put research participants, the public or vulnerable groups at **risk**, e.g. spatial data that would reveal the location of an endangered species can be justifiably withheld to protect the species from poachers

Notes:

It should always be made clear to future users which variables have been redacted, aggregated or anonymised in a datasets and why.

¹ <http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/RCUKCommonPrinciplesonDataPolicy.pdf>

3.2 Contractual obligations

- The **copyright** to the data, or parts of the data, is not owned by the University, and the copyright holder does not permit data sharing (e.g. via SHURDA), or only under certain conditions
- The research was conducted under a contract with a funder that states that the research data must remain **confidential** (for a certain number of years)
- If any **existing datasets** obtained from third parties have been used, the terms under which these data can be used and shared may prohibit or limit sharing. This may include datasets from online repositories and data shared by project collaborators

Notes:

When research is funded by an external partner, or where an external partner makes a contribution to a project, the partner may be awarded Intellectual Property Rights in the results including research data. This may mean that the results must be kept confidential by the University and can only be released under certain conditions. Collaboration agreements should address the governance of storing, accessing and publishing primary research data.

The data availability statement to be inserted in any published work should, as far as possible, provide details of where the data can be obtained, so that other researchers can consult or request access to the same data. This is particularly important for publicly available data.

The terms under which the research was paid for may stipulate admissible embargo periods.

When copyright or other permissions have not yet been cleared but may be cleared in the future, an embargo could be applied.

3.3 Commercial restrictions

- The data is **commercially sensitive**. These data could be provided by commercial partners, possibly under terms of a collaboration agreement, or the data could be generated, for example obtained in interviews with participants employed by external organisations (which would require informed consent)
- The data might have **commercial potential**. These data might eventually be suitable for sharing, but public access can be justifiably delayed to allow time to assess and protect the commercial potential of the research, e.g. applying an embargo while patent applications are filed. Not all funders and publishers accept restricted access to protect patent applications

Notes:

Every endeavour should be taken to maximise the potential for data sharing. It may be possible to make commercially sensitive data available only to certain users, such as bona fide researchers in a research organisation, and for a certain purpose, such as to verify and comment on a publication — in which case a non-disclosure agreement will be required. In other cases, an embargo period may be a good solution. It is essential to always check the relevant funder's and publisher's policies and any other agreements established with research partners.

3.4 The right to reasonable first exclusive use

- The *Draft Concordat on Open Research Data* states that 'researchers who generate original data must have **reasonable right of first use for a short and well-defined period**, which may vary by subject and disciplinary area. [...] In some circumstances, the right of first use could include the withholding of initial datasets until later related datasets have been developed. This

could be justifiable if such an action were to advance conceptual understanding of central concepts carrying implications for the research field, or in studies that address long-term changes or developments. [...] [I]n all cases data supporting and underlying publications should be accessible by the publication date in citeable form²

- RCUK's *Guidance on best practice in the management of research data* addresses 6 valid reasons for a period of exclusive use:³
 - pending publication
 - effort required to produce data
 - data preparation
 - on-going data collection
 - protection of intellectual property
 - recognition of intellectual contribution

RCUK nevertheless states that as a general rule, research data should be made publicly available no later than the publication of the main findings.

4 More information

- <http://research.shu.ac.uk/rdm/legal-and-ethical-aspects.html>
- <http://ukdataservice.ac.uk/manage-data/legal-ethical/obligations>
- <http://www.bath.ac.uk/research/data/sharing-reuse/restrictingaccess.html>
- <http://www.rcuk.ac.uk/research/opendata/>
- <http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/RCUKCommonPrinciplesonDataPolicy.pdf>

² <http://www.rcuk.ac.uk/research/opendata/> (September 2015)

³ <http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/RCUKCommonPrinciplesonDataPolicy.pdf> (July 2015)