

Anthropology research data management and sharing guidance

Version 1	29/05/2019	Eva Borger	Finalised document following review by researchers, approved by the HoS and HoDept
-----------	------------	------------	--

Contents

Introduction	1
Scope.....	2
What is data	2
Data sharing, data protection and ethics.....	2
How and where to share.....	3
Depositing in Pure.....	3
Secondary data	4
Data management plan (DMP)	4
Data storage options to consider.....	5
Data access statements	5
Questions about this guidance	5

Introduction

Funders have introduced open data policies¹ that require researchers to share their data. Specifically, all UK Research Councils, The Wellcome Trust, Scottish Funding Council and others subscribe to the UK Research and Innovation (UKRI) [Concordat on Open Research Data](#) and typically also require the submission of a data management plan. The University's Research Data Management policy fully incorporates the Concordat's principles². In addition, the University's ethical application processes now requires detailed data management statements. However, these policies often lack a clear definition of 'data' in different scholarly disciplines and guidance on how researchers can comply in practice.

¹ <https://www.st-andrews.ac.uk/library/services/researchsupport/researchdata/researchdatapolicies/fundersresearchdatapolicies/>

² <https://www.st-andrews.ac.uk/staff/policy/research/researchdata/>

Therefore, the Research Data Management team are working closely with the Schools to develop discipline specific guidance with the aim of answering the following questions:

- What are 'data' or 'digital objects'? When does a research result or output become data?
- What data should be deposited or registered in Pure to be compliant with funder policies?
- What is a 'theoretical' paper where data sharing does not apply?
- What exceptions to data sharing are acceptable?

Scope

The discipline specific guidance aims to offer a reference document for researchers considering sharing of data and materials underpinning their research as part of the publication process or during data management planning. Its development has been a collaborative process between the RDM team and researchers at St Andrews. **This document is not intended as a policy, but rather as guidance agreed by researchers within the department, which will be reviewed regularly.**

This document is also *not* designed as a data storage guide. Whenever possible, research data should be stored and backed up on University infrastructure, such as the University network storage or Office365 Cloud storage³. These can, under certain circumstances, also be used for confidential data in line with the University's Information Classification and research data protection guidance⁴. The Research Data Management team advise to back up any research data on the University network storage periodically during a project and archive any data on the network at the end of the project. Data that can be made publicly available should then be deposited in a public repository or archive.

What is data

Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements. - CONCORDAT ON OPEN RESEARCH DATA

It can also be conceptualised as *the valuable information a researcher or others might want to 'come back to' to build on at a later stage.* In Anthropology, this might comprise:

- transcripts or statements (digital or analogue)
- photographs
- videos
- quantitative information or measurements
- field notes
- new data produced from secondary data sources

Data sharing, data protection and ethics

Not all data are shareable. Especially in the field of Anthropology, data must often be classified as confidential because of ethical or legal considerations arising from the nature of the data itself or the setting in which it has been collected. This might mean that from some studies, there might not be any shareable data/ digital outputs at all, while in other instances, subsets of the data or anonymised versions might be shareable.

³ <https://www.st-andrews.ac.uk/library/services/researchsupport/researchdata/datastorage/datastorageoptions/>

⁴ <https://www.st-andrews.ac.uk/itsupport/security/classification/> and <https://www.st-andrews.ac.uk/terms/data-protection/>

Non-availability of research data for these reasons is acceptable and does not violate any institutional, funder, or journal/publisher data sharing policies. However, such restrictions should be explained in data management plans and in a data access statement in publications.

What to do: 1) Consider if it is possible to extract valuable, shareable information for digital preservation or future re-use. Where necessary, highlight necessary restrictions to data sharing in your management plan. 2) **Fulfil all obligations on data handling and security resulting from ethical review or legal agreements**⁵. 3) Include a statement in your publication that clarifies restrictions and instructions for potential conditional access.

How and where to share

The course of action for the three most common categories of data are summarised in Figure 1.

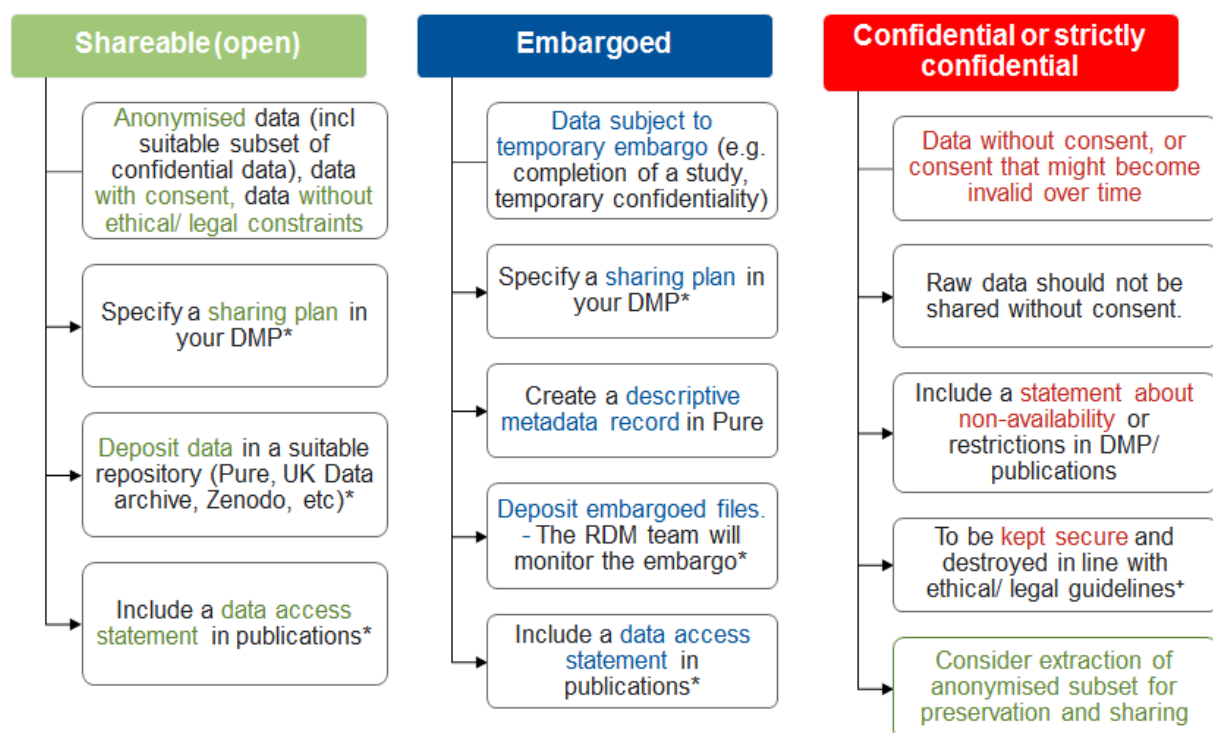


Figure 1: Options for different types of data: *Advice is available from the RDM team at research-data@st-andrews.ac.uk;

*Advice is available from your School ethics convenor or dataprot@st-andrews.ac.uk

Depositing in Pure

As a general rule, whenever there is a shareable digital or physical output, whether this can be made publicly available without restrictions, requires an embargo or restricted access, this should be registered as a 'dataset' record in Pure either as:

1. a **public dataset record** with digital files attached which are made available on the Research Portal at <https://risweb.st-andrews.ac.uk/portal/>
2. an **embargoed dataset record** with embargoed digital files attached, for which the metadata and embargo details are made available on the Research Portal

⁵ For advice, contact your School Ethics Convenor, Stavroula Pipyrrou, at sp78@st-andrews.ac.uk

3. a **reference record** that points to another repository or database⁶
4. a **metadata record** that describes conditions and instructions for restricted access.
5. a metadata record that describes instructions for access to physical resources.

Guidance on how to create a dataset record in Pure can be found at <https://www.st-andrews.ac.uk/research/support/pure/using/>.

Secondary data

Secondary data obtained from other sources, e.g. databases, archives, does not need to be (and should not be) deposited again.

What to do: 1) Cite secondary data sources⁷ in any publications as appropriate, in addition to citing associated reports or publications. 2) Create a dataset deposit for any **new derived data** in Pure or other repository, adhering to any licensing conditions associated with the original data source. 3) Where new derived data has been deposited in Pure, include a data access statement with the DOI for the dataset record in your publications.

Data management plan (DMP)

Most funders now require the submission of a data management plan (DMP) alongside grant applications or ask specific questions on data management. To complement this, the University's Research Data Management policy also strongly encourages researchers to create a data management plan for all of their projects irrespective of funder requirements⁸. In addition, providing details on data management is also a substantial part of the University's ethical application process that can be fulfilled through submission of a DMP. A DMP provides information about the data created or used during a research project. While there are variations between templates provided by funders or the questions asked in an application, a DMP essentially serves to establish:

- What kind of data will accrue?
- How much data will accrue? - That is, will more data storage need to be purchased that could be costed in a grant application?
- Which data security requirements will need to be met (based on existing ethical and risk assessments)?
- Where will data be stored during the project and for long-term preservation?
- Who will be responsible for handling, managing storage of the data and for access control?
- Can all or some of the data be shared? If so, when and how will the data be shared? If not so, why not?

By addressing these points, a DMP summarises the **technical aspects** of handling, storing and managing research data, any **costs** that might be associated with this (and be claimed as part of a grant application), as well as considerations regarding **opportunities for, or restrictions to, data sharing**.

⁶ Data or materials might available through a subject-specific repository or database (e.g. <https://www.re3data.org/search?query=anthropology>) through a collaborator's institutional repository.

⁷ <https://www.st-andrews.ac.uk/library/services/researchsupport/researchdata/publisharchiveandpreserve/datacitationandaccessstatements/>

⁸ <https://www.st-andrews.ac.uk/staff/policy/research/researchdata/>

Guidance available on the University's Research Data Management web pages, at <https://www.st-andrews.ac.uk/library/services/researchsupport/researchdata/planningfordata/datamanagementplan/> may help to answer the questions above. In addition, the Research Data Management team can provide advice on completing a DMP tailored to each project and specific funders.

Data storage options to consider

Each data management plan should address the handling and storage of research data during and beyond a project⁹. Ideally, all research data should be stored on University infrastructure (e.g. university owned machines, local network storage or Office365), so the plan should outline how this will be achieved and who is responsible for overseeing this and managing access. In doing so, special consideration should be given to the suitability of different storage locations and, data transfer and data access arrangements with respect to data protection and privacy legislation such as GDPR¹⁰.

Data access statements

In line with funder policies and good practice, all publications, including PhD theses, should contain a statement on the availability of any digital outputs or underlying research data. Scenarios, which such statements could describe, include:

- publicly shareable data are available in a data repository at a public URL
- publicly shareable data exist, but are embargoed for a period of time and will be available from a repository after the embargo has expired
- shareable data exist, but can only be made available through controlled access
- data produced or used during the study are not publicly shareable for legal or ethical reasons
- no new digital outputs/ research data were produced

The Research Data Management team issues digital object identifiers (DOIs, permanent links) for all data deposited in Pure, which should be used for data citation where a dataset record is available at present or where will become available in the future, subject to embargo.

The precise content and formatting of data access statements are not prescribed. However, examples are provided on the Research Data Management web pages: <https://www.st-andrews.ac.uk/library/services/researchsupport/researchdata/publisharchiveandpreserve/datacitationandaccessstatements/>.

Questions about this guidance

Any questions about sharing or publishing research data or the guidance provided above can be directed to the Research Data Management team:

Old Union Diner, Butts Wynd
St Andrews, KY16 9AL, Fife, Scotland
Tel: +44 (0)1334 462322 or +44 (0)1334 462343
Email: research-data@st-andrews.ac.uk

⁹ See: <https://www.st-andrews.ac.uk/library/services/researchsupport/researchdata/datastorage/datastorageoptions/>

¹⁰ See <https://www.st-andrews.ac.uk/terms/data-protection/> for guidance on how to comply with data protection legislation or contact dataprot@st-andrews.ac.uk.