

## Guidelines for selecting research data for retention and preservation

*The individual researcher is responsible for the retention and archiving of data and must comply with any external requirements (e.g. funders), and the terms on which ethical approval was granted.*

*Where there are no specific external requirements for retention, the researcher should keep the data as long as is necessary for the purpose of the research, and in line with any data collection agreements, or funder or institutional requirements.*

[Code of Practice for Research at the Open University](#), July 2013

### Contents

1. Introduction [“What are these guidelines for?”](#)
2. Data selection [“Why not keep everything?”](#)
3. Data retention periods
4. Expired retention period [“The retention period’s over, what should I do?”](#)
5. [Links to related Open University Resources](#)
6. [Links to other useful resources](#)

### 1. Introduction

#### **“What are these guidelines for?”**

Research data will typically be stored, managed and often shared throughout the course of a research project. After a project has ended, the data will need to be retained for a specified period. When considering longer term archiving, it can be tempting to keep it all data for possible future use. However, not all data can or should be preserved. Evaluating your data will enable you to decide which data to deposit in a trusted archive or repository.

This set of guidelines are a starting point for determining how long data should be retained and selecting which data should be preserved in the long term. They may help you at the beginning of your project when you are composing your data management plan. Including information on the preservation of research data will help you to clarify what the requirements for your data are and enable advance planning for long-term storage of your research data. They may also be useful towards the end of your project, when you are sorting your data for deposit in a data archive.

**Please note that this is a guidance document and reference should also be made to:**

- The Open University’s [Retention Schedule](#)
- Your funder’s specific policy (consult this [Overview of Funders’ Requirements](#))
- The Open University [Ethics Principles for Research involving Human Participants](#)

## 2. Selecting data for preservation

Although it may be tempting to deposit all of your data into an archive or repository, just in case it proves useful in the future, this is not always possible, as archiving everything could prove to be costly and time-consuming and in some cases, unethical.

While it is true that the cost of data storage has declined over time, the cost of organising, describing and then maintaining data in a usable form has a considerable cost. In fact, the cost of data “curation” is often much greater than the cost of data storage.

If you are gathering data from human research participants, you will also need to ensure that you have gained their informed and valid consent for the specific archival, share and re-use management plan for the data, and you will need to have a plan to securely dispose of any data which cannot be shared.

### *“How do I know what to deposit?”*

#### **Evaluation criteria**

The following questions may help you when considering which data to deposit:

- **Are there any ethical issues to consider?**  
Have you gained consent from your research participants to archive, share or reuse the data?
- **Do the data have special scientific or historical value?**  
Does evidence of current research in your field suggest that your data will be important in the future?
- **Are the data unique?**  
Would the information derived from your dataset be at risk if the dataset were lost?
- **Do the data have a high re-use potential?**  
Are the data likely to be of broad interest? Has their reliability been assured?
- **Can the data be easily reproduced?**  
Would it be feasible to replicate the data? Would it be financially viable?
- **Is there a strong economic case for preservation?**  
Are the estimated costs related with data curation justifiable when you consider the potential future benefit?
- **Are the data in support of a patent application?**  
It is necessary to retain data in support of patent applications because we may be called upon to defend our patents in court and the original research data can be critical in this process.

#### **Policies and requirements impacting on data selection**

The following policies and formal requirements will also impact on your decision:

- **Research funder policy**  
Most major research funders have guidelines or requirements regarding which data should be retained and shared after a research project is concluded. This [Overview of Funders' Requirements](#) gives details of the requirements of the main funders of research projects at the OU (please note; these policies are liable to change, so reference should also be made to the funders' own policies).
- **Data centre policy**  
If you are going to deposit your data into a subject-based data centre, subject-specific evaluation criteria may apply. Where this is the case you should follow the guidance provided by the data centre in question.
- **Academic publisher requirements**  
Increasingly, academic publishers also require data which underpins a publication to be retained and shared. Check with your publisher if this is the case, as they may have requirements for where it is stored and for how long.
- **Intellectual Property Policy**

### ***“This seems complicated. How can I be sure I’ve deposited the right data?”***

The decision of what data to select for long-term preservation will never be entirely objective, as it is impossible to know exactly what information will be useful in the future. However, by thinking it through carefully, abiding by any relevant policies or requirements (e.g. from funders) and documenting the decisions made and the reasons for them, you can ensure that your selection process will be as fair and as accurate as possible.

It is helpful to consider the long-term preservation of your research data when writing your Data Management Plan; by planning which data to archive and incorporating a research data management strategy into the life-cycle of your project you will find the process of selecting and preparing data for long-term preservation easier and quicker.

### ***“What about the data and research records I decide not to deposit?”***

The Open University’s Retention Schedule advises that all research records (which include raw and analysed research data) be kept for ten years from publication. Any research records or data which you decide not to deposit in an archive or repository should still be kept for ten years however, in these cases, the management and accessibility of those data and records remains your responsibility.

## **3. Data retention periods**

### ***“How long should I keep my data?”***

#### **Open University requirements**

The Open University’s Retention Schedule advises that research records (which include raw and analysed research data) be kept for ten years from publication. After ten years, if the data prove to still be pertinent, then a longer retention period may be considered.

In the case of data retained in support of patent applications, the data must be retained for the lifetime of the patent (up to 20 years or until the patent is lapsed). Please discuss any queries on data related to patents with RSQ.

For more information, consult the University’s [Retention Schedule](#)

#### **Funder requirements**

Different funders have different requirements about where, when and for how long data should be kept post-project. This [Overview of Funders’ Requirements](#) gives details of the requirements of the main funders of research projects at the OU (please note; these policies are liable to change, so reference should also be made to the funders’ own policies).

#### **Retention “in perpetuity”**

Certain types of data might have a recommendation that retention is ‘in perpetuity’. Usually this relates to data that cannot be recreated (e.g. observations of unique events). Although true for all retained data, it is important to ensure that additional care is taken to hold this data in a durable format.

#### **Funders with differing retention periods**

If the retention period differs between your funder and the University, data should be retained for the longer of the periods.

#### **Data and publication**

Increasingly, publishers are requiring the deposit of supporting data with the article, while others require the provision of a link to the data. You should consider this when deciding how long you will need to retain the data and it may influence your choice of storage location.

#### 4. Expired retention period

##### *“The retention period’s over. What should I do?”*

As part of the deposit process you will be asked to consider what should happen at the end of the retention period and who is responsible for carrying this out.

At the end of the retention period you may need to re-evaluate your data in case it needs to be retained for longer than you had planned. You will need to consult the data centre or archive in which it is stored in order to arrange an extension.

If you leave the OU before the end of the retention period, it is your responsibility to hand over custody of the research data to a colleague who will be able to make any future decisions about it.

For more information on securely disposing of your data consult the **Media Disposal Policy** within the [Information Security Specific Policy Set](#)

#### 5. Links to related Open University resources

Open University’s [Retention Schedule](#)

[Overview of Funders’ Requirements](#)

[Code of Practice for Research at the Open University](#)

[Intellectual Property Policy](#)

The Open University [Ethics Principles for Research involving Human Participants](#)

#### 6. Links to other useful resources

[JISC’s HEI Records Management Policy](#) (2007)

DCC [Overview of Funders’ Data Policies](#)

Wendy Mears

6<sup>th</sup> October 2016