

Pseudonymisation for Research

Introduction

Pseudonymisation (also known as codification and other names) is the process of providing research participants with 'ID Numbers' or Pseudonyms in order to protect their identity.

When do you start the pseudonymisation process?

As soon as you start collecting your participant names, contact details, etc. you should start entering these in your pseudonymisation spreadsheet / database and assigning them a Pseudonym or Participant ID number / name.

This Pseudonym will then be used in place of their name in your research instruments when you start gathering your data e.g. in file names and transcripts.

This means that the Pseudonymisation Spreadsheet is the 'key' for connecting the participant to the survey instrument and this must only be possible by those researchers with the password to the Pseudonymisation Spreadsheet. You should not be able to identify your participants directly from the research instruments. This is easier in surveys than interviews, for example, but you should be mindful to protect the personal data of your participants at all times.

Pseudonymisation Spreadsheet/s for Research (examples)

There are much more technical ways of pseudonymising personal data – this is a simple way of doing it.

These spreadsheets must be password protected and kept separate to the research data on the X or V Drive that the project has been allocated. The password should only be shared with other colleagues who specifically need access to the personal data e.g. access is restricted.

In the password protection spreadsheet (pseudonymisation master/key document/database), we'd expect something along these lines:

Pseudonym	Name	Contact details	Job Title	Organisation	Etc	Etc
Participant 001	John Doe	j.doe@acmemuseums.co.uk	Curator	Acme Museum		
Participant 002	Arthur Smith	arthurs@gothamcity.org.uk	Gallery Manager	Gotham City Museum		
Participant 003						

No special category personal data should be held in the master pseudonymisation database, but demographic data and any special category personal data should be held in a separate password protected spread sheet, along these lines:

Pseudonym	Age	Postcode	Gender	Ethnicity	Etc	Etc
Participant 001	25-35 years	EH4 1Y	M	White British		

Participant 002	18-24 years	EH11 4B	M	Black - other		
Participant 003						

Please remember that these are just examples, and the data collected must be only that which is absolutely necessary to achieve the research aims and objectives.

Your research instruments (surveys, questionnaires, interviews, etc.) would all then be pseudonymised accordingly, e.g file names provided below, but the same principle would be used within the documents/files:

Interviews:

Interview-Participant001
Interview-Participant002
Etc.

Surveys:

Survey-Participant001
Survey-Participant002
Etc.

The pseudonymisation spreadsheet needs to be kept as long as is necessary for your research project, ideally until any audit (or challenge) which requires verification of the findings has been completed (this may include peer reviews or similar – any event which may require evidence to be produced to verify the findings). You'll need to think on a practical level of how long it's necessary for you to keep the data e.g. how long is it actually required for the project, and then if you will need it for anything else in connection with the project. A rough guide is that it will be destroyed once any audit to verify the findings has been conducted, usually within 12 months of the project finishing. If there are funder or other similar requirements it might be possible to keep it for up to 6 years, as you would for records documenting research ethics consent. There are exceptions to this e.g. for large scale funded projects and long term studies, so please consider your actual requirements.