

EASHW TIPS – ANONYMITY AND PSEUDONYMITY

Every researcher guarantees confidentiality and strives for anonymity.

4. RECOGNISABILITY: IS IDENTIFICATION OR RE-IDENTIFICATION OF PARTICIPANTS POSSIBLE IN THE ANALYSIS DATA AND/OR RESULTS?

There is a difference between (non-)anonymous research and (non-)anonymous data. There is also a difference between non-anonymous, pseudonymous and anonymous data. These differences are important.

For the difference between anonymity and pseudonymity, we start from the possible recognisability of participants. We also use the terms *identification (participants are recognizable)* and *re-identification (participants can be made recognizable again)*.

ANONYMOUS & NON-ANONYMOUS RESEARCH

Research is anonymous if "no one is recognizable by anyone - with the use of reasonable resources"

- Even if only one person can recognize a participant, the study is not anonymous. This means that as soon as even one researcher can recognize a participant, there can be no question of anonymous research. Here too, the boundary is not entirely black and white: street research, such as conducting a quick survey, can be considered anonymous if the researcher does not know participants in these surveys.
- '*With the use of reasonable resources*': all those who can trace individuals in data files with special equipment or actions are not included here.

Research is NOT anonymous when (list is not exhaustive):

- Researchers come into contact with participants.
- Recordings are made in which the face and/or voice are present (use of photos, audio or video recordings).
- Recordings are collected in which the face and/or voice are present (e.g. photovoice studies)
- Contact details are requested.
- (Other) identity data are requested.
- Data or combination of data can lead to the recognition of a participant.

NON-ANONYMOUS, ANONYMOUS AND PSEUDONYMOUS DATA

Data is **non-anonymous** if someone **can recognize individuals directly or indirectly in the dataset**.

This is the case, among other things, for the following types of data:

- *Direct recognisability*: Names, contact details, audio or visual material and the like.
- *Indirect recognisability*: based on the combination of some data. For example, the name of a company and the position of an employee, or the age of an *outlier* in a population, etc. We ask you to estimate whether participants can be recognizable based on the data you collect.
- When in doubt, consider the data to be non-anonymous.

Data is **anonymous** if **no one can directly or indirectly recognize anyone, nor can they re-identify**.

This is the case if:

- You have never collected data with which one can directly or indirectly recognize someone, OR

- *You anonymize your data:* You have collected data with which participants can be recognized, but you permanently delete this data. (For example, destroying audio recordings after transcription where participants are no longer recognizable in the transcripts).

Data is **pseudonymous** if one **can re-identify persons**.

This is the case if *you pseudonymise data* by working with:

1. A database in which participants are recognizable,
2. A data file in which participants are made unrecognizable with a code, and
3. A key file that can link the code from file 2 to the data from file 1.

EASHW ADVISES:

- Prefer anonymous data collection if this is possible,
- Prefer the anonymization of data to pseudonymization,
- Use the terms correctly and consistently throughout your application,
- Inform your participants correctly:
 - Is your data collection anonymous or not?
 - If not: will you anonymize, pseudonymise or keep data non-anonymous? Explain why you choose which option and from which point data may or may not become anonymous/pseudonymous.

VISUAL – PSEUDONYMIZING DATA VS. ANONYMIZING

This is a visual representation based on a fictional example:

DATASET

Full dataset
 Name
 Email
 Employer + position
 Gender
 Nationality
 Average number of steps per day
 Average number of hours of sleep
 Average number of publications per year

Data with which participants are recognizable:
 Name
 Email
 Employer + position

Data with which participants are not recognizable
 Gender
 Nationality
 Average number of steps per day
 Average number of hours of slack
 Average number of publications per

PSEUDONYMIZATION =

Data with which participants are not recognizable
CODE
 Gender
 Nationality
 Average number of steps per day
 Average number of hours of slack
 Average number of publications per



Data with which participants are recognizable (KEY file):
CODE
 Name
 Email
 Employer + position



Full dataset
 Name
 Email
 Employer + position
 Gender
 Nationality
 Average number of steps per day
 Average number of hours of slack
 Average number of publications per

Saved to protected file on OneDrive
 May be shared

KEY FILE
 Saved to SEPARATE protected file on OneDrive
 May NOT be shared

Re-identification is possible

ANONIMIZE =

Data with which participants are not recognizable
CODE
 Gender
 Nationality
 Average number of steps per day
 Average number of hours of slack
 Average number of publications per

Data with which participants are recognizable:


Full dataset
 Gender
 Nationality
 Average number of steps per day
 Average number of hours of slack
 Average number of publications per year

Saved to protected file on OneDrive
 May be shared

PERMANENTLY DELETED
 (no copy kept anywhere)

You only keep an anonymous dataset