KT4D Data Protection

Pseudonymisation Guidelines for Use Cases

1. General Data Protection Considerations to keep in mind

How should I take notes during the Use Cases?

Please proceed to take notes on paper or electronically as you would normally do. You can go into as much detail as needed in, e.g., the case of ethnography or participant observation. These details might be relevant for your research.

Ensure that your notes:

a. Are kept in a safe place (don't leave your notebook or computer unattended, ensure your computer has a password and if possible, take the notes directly in a drive that allows you to ensure that the file is not lost in the case of theft or damage to your device).
b. Are only accessible to the people who strictly need to access them (don't share them in a folder that is accessible to your entire organization, keep them protected by a password).

2. What else should I take into consideration from the data protection perspective?

a. BE CAREFUL TAKING PHOTOS

If you are taking photos for promotional purposes that will be shared online, make sure that you have collected the specific (so for the photos, not for the research in general) free and informed consent of individuals to use their photos for these purposes, **even if you are not going to capture their face**. Inform them that if they do not wish to have their photos taken that they won't suffer any negative consequences and be sure to not take their photo!

This is because it may still be possible to identify people even without their face being captured in the photo and vulnerable data subjects may be involved in the Use Cases.

a. COLLECT INFORMED CONSENT

Make sure that your information notice for research is clear and that consent is collected. Mention to the participants that you will pseudonymise their data – do not incorrectly refer to anonymisation.

b. REMEMBER THE DEFINITION OF PSEUDONYMIZATION

Pseudonymization is the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person (Art. 4(5) GDPR).

HOW SHOULD I PSEUDONYMISE?

STEP 1: CODE THE DIRECTLY IDENTIFIABLE DATA IN THE DATASET USING THE COUNTER METHOD

Your goal is to code all directly identifiable information.

Use pseudonyms and generic descriptions to pseudonymise textual information, e.g., in interview transcripts, and indicate replacements with [square brackets], e.g. [Person 1] works for [a financial organisation] in [country].

Directly identifiable data includes any information that could directly lead to the identification or singling out of a single person, e.g.:

- A name and surname; a home address; an email address such as name.surname@company.com; an identification card number; PNR number; passport number; an Internet Protocol (IP) address; a cookie ID; the advertising identifier of your phone; data held by a hospital or doctor; a description of someone with multiple identifiers (see example from ICO below).
    - If you are able to single out/distinguish one individual from another individual by looking solely at the information you have in front of you, that individual will be identified (or identifiable).
    - Remember, it’s not just about names. A combination of identifiers may be sufficient to identify the individual.
    - Examples from ICO[1] of how data can be combined to identify someone:
        - The elderly man who lives at number 15 Purple Street and drives a Porsche Cayenne.
        - Megan Smith’s foster mum, from Year 4 at Broomfield Junior School.
        - A description of an individual may be personal data if it is processed in connection with a neighborhood watch scheme, for

[1] Can we identify an individual directly from the information we have? | ICO

the purpose of identifying an individual as a potential witness to an incident.

- If an individual is directly identifiable from the information in the dataset or video, this may constitute personal data.

Use the "Counter method" to code people.

- ENISA describes the "Counter method" as "the simplest pseudonymisation function, where the identifiers are substituted by a number chosen by a monotonic counter. Its advantages rest with its simplicity, which make it a good candidate for small and not complex datasets. It provides for pseudonyms with no connection to the initial identifiers (although the sequential character of the counter can still provide information on the order of the data within a dataset). However, the solution may have implementation and scalability issues in cases of large and more sophisticated datasets."
  - Concretely, you could pseudonymise as exemplified in this invented text:
    - Karen Kranston, UK resident, stated that she felt excluded from democratic processes in her home country of Yellowland due to the fact that people from the Purple tribe, to which she belongs, are considered to be inferior by the Yellow majority. However, in the UK she feels that she can actively participate.
    - [Person 1], [Country 1] resident, stated that she felt excluded from democratic processes in her home country of [Country 2] due to the fact that people from the [Tribe 1] tribe, to which she belongs, are considered to be inferior by the [Tribe 2] majority. However, in [Country 1] she feels that she can actively participate.

STEP 2: ENCRYPT THE CODING KEY AND KEEP IT SEPARATELY FROM THE CODED DATA FILE

Don't keep the coding key (i.e., Country 1 = UK) under your control, have someone else keep it (use the "four-eye-principle"). The same logic is applicable to recordings.

STEP 3: BLUR FACES IN VIDEO AND DISTORT AUDIO

Blur faces and distort voices, e.g., blur faces using photoshop, OpenShot, change pitch of voices by using tools such as Audacity.

STEP 4: DOUBLE CHECK YOU HAVE REMOVED ALL DIRECTLY IDENTIFIABLE DATA FROM THE DATASET

Your objective with coding is to ensure that the dataset cannot be used by anyone else to identify the data subjects using reasonable means which are likely to be used.

STEP 5: ONLY SHARE PSEUDONYMIsED DATA WITHIN THE CONSORTIUM

STEP 6: UPDATE YOUR RECORD OF PROCESSING ACTIVITIES (AND KT4D DATASHEET) WITH THE PSEUDONYMIsATION PROCESSING ACTIVITY

Pseudonymisation is a data security measure which can help you comply with your obligations under Arts. 5, 25, and 32 GDPR. However, the process of pseudonymising data constitutes data processing. For this reason, the activity must be logged in your Art. 30 Record of Processing Activities.