Anonymisation plan

# General information

|  |  |
| --- | --- |
| **Dataset name** | What is the name of the dataset? |
| **Creator(s) of the plan** | Who has contributed to the writing of the anonymisation plan? |
| **Responsible person(s)** | Who is involved in the anonymisation process? |
| **Date / version number** | What is the creation date and/or version number of this version of the anonymisation plan? |
| **Project summary** | Describe the research project in a short summary and provide contextual information about the project. |
| **Related sources** | Is it possible to connect the information in the data to other sources of information? This can be research data about the same target population available elsewhere, publicly available information, local knowledge or personal information about other people. |

# Population and sampling

|  |  |
| --- | --- |
| **Population** | Describe the target population of the study, e.g. demographics, geographical location, distribution of traits within the population. The size of the population is important, because the smaller the population or the phenomenon under observation, the greater the possibility of being able to identify an individual person. |
| **Rare phenomenon** | Do individuals belonging to the population share a rare phenomenon? This can for example be an unusual event, highly specific job title or very rare disease. Rare or exceptional traits tend to increase the risk for identification. This can also be a skewed population distributions with highly identifiable outliers (e.g. age, gender, ethnicity). |
| **Sampling** | How is the sampling conducted? How many people belonging to the population were included in the sample? You should consider how random belonging to the population and being selected to the sample are. |

# Dataset age

|  |  |
| --- | --- |
| **Age** | When was the data in the dataset collected? The older the data, the more difficult it is to identify individuals, because the information changes over time. |
| **Change over time** | Have the records in the dataset changed over time since they were first collected? In other words, how accurate and up-to-date is the information in the dataset? |

# Information value

|  |  |
| --- | --- |
| **Significant information** | Which information in the data is essential with regards to your research? What information must be preserved during anonymisation? Complete removal of values with identifiers should only be done with less significant information. |
| **Irrelevant information** | Which information in the data is irrelevant with regards to your research? What information can be removed during anonymisation? |

# Identifiers in the dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Identifier** | **Di[[1]](#footnote-1)** | **In[[2]](#footnote-2)** | **Se[[3]](#footnote-3)** | **Decision whether to anonymise, chosen anonymisation technique and rationale[[4]](#footnote-4)** |
| Age |  |  |  | E.g. Generalised to 10-year categories, suppressed outliers of 85 years and older. This allows me to identify different age groups in the dataset, which is necessary to answer my first research question. By removing outliers there is less risk for accidental identification. |
| Identifier |  |  |  | Decision, anonymisation technique and justification. |
| Identifier |  |  |  | Decision, anonymisation technique and justification. |
| Identifier |  |  |  | Decision, anonymisation technique and justification. |
| Identifier |  |  |  | Decision, anonymisation technique and justification. |
| Identifier |  |  |  | Decision, anonymisation technique and justification. |
| Identifier |  |  |  | Decision, anonymisation technique and justification. |
| Identifier |  |  |  | Add extra rows as needed. |

# Other identifiable information

|  |  |
| --- | --- |
| **Open-ended responses** | How are open-ended responses processed? |
| **Combinations** | Which combinations of information in the data could be used to identify an individual? |
| **Third persons** | Does the dataset contain information related to third persons? Can individuals be identified based on this information? How are these identifiers processed? |
| **Future risks** | Could there be a remaining possibility that individuals can be identified in the data, now or in the future? If yes, when should the anonymity of the data be reviewed again? |

1. Direct identifiers refer to Information that is sufficient on its own to identify an individual includes a person's full name, social security number, email address containing the personal name, and biometric identifiers (fingerprints, facial image, voice patterns, iris scan, hand geometry or manual signature). [↑](#footnote-ref-1)
2. Indirect identifiers(or quasi-identifiers) are the kind of information that on their own are not enough to identify someone but, when linked with other available information, could be used to deduce the identity of a person. Background variables and indirect identifiers include, for instance, age, gender, education, status in employment, economic activity and occupational status, socio-economic status, household composition, income, marital status, mother tongue, ethnic background and place of work or study. Indirect identifiers relating to region of residence include, for example, postal code, neighbourhood, municipality, and major region. Date can also be an indirect identifier. Date of birth is the most common example, but dates of death and dates of newsworthy events may also be indirect identifiers in research data when combined with other information. [↑](#footnote-ref-2)
3. Sensitive identifiers contain categories of personal data specified in data protection legislation concerning racial or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, data concerning health, sexual life or sexual orientation, or genetic or biometric data for identifying a natural person. Other information may also be sensitive by nature. Sensitivity can be assessed by, for instance, considering whether the phenomenon is taboo or how much the disclosure of the information could provide harm to a person, organisation, or other unit of observation. [↑](#footnote-ref-3)
4. Describe if you want to anonymise this identifier, what anonymisation technique you want to use and why this is a necessary and suitable anonymisation technique. [↑](#footnote-ref-4)