

Identifiability and Data Management

Managing and protecting research data

Identifiability of research data: provide a clear outline of the identifiability of collected/stored data

Data Type	Definition	Example
Identifiable data	contains information which identifies specific participants	includes direct identifiers such as participants' names or indirect identifiers allowing identification of participants through a combination of indirect identifiers such as date of birth and postal code.
Coded data	does not contain direct identifiers; identifiers are replaced with a code which can re-identify participants	instead of labeling the data using the participant's name, the researcher labels it using a code such as "P1"
Anonymized or de-identified data	irrevocably stripped of direct identifiers (no code is kept); low risk of re-identifying participants from remaining indirect identifiers	the researcher separates the demographic information from survey responses and deletes the demographic information so that the survey answers cannot identify specific participants
Anonymous data	never had identifiers and risk of identifying participants is low	an online survey which does not collect demographic information, IP addresses, or any other identifiers

Data management plan: detailed description of data collection, access, storage, and destruction

Data collection methods: tools, platforms, instruments, or software used

Data storage locations: secure servers, cloud storage, or password-protected local PC

Access controls: who will have access and how access is managed (passwords, encryption, physical controls, etc.)

Confidentiality strategies: whether data will be anonymous, anonymized/de-identified, or coded, when identifiers will be removed, how you will keep participants from being identified and whether/how you will use AI tools

Retention and destruction timelines: duration of data storage, deletion methods, whether data may be archived or shared (and under what conditions)