Data protection and data processing in scientific studies.

Joachim Westenhöfer, 08.11.2022, adopted by the Ethics Committee of the University of Applied Sciences Hamburg 08.11.2022

# Preface

This text provides a short overview of the major consequences of data protection regulations and good scientific practice and is intended to provide assistance for reseachers.

# General terms

Regarding data, we have to distinguish two or three types of data. These terms are defined in note (26) of European General Data Protection Regulation (GDPR: Regulation (EU) 2016/679):

**Personal data**: any information concerning an identified or identifiable natural person

**Pseudonymised data**: Personal data which has undergone pseudonymisation.

**Anonymised data**: “information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable”.

Important note: Data protection regulation and rules do not apply to anonymised data but applies to personal data and pseudonymised data.

Pseudonymised data, that is “personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.” (General Data Protection Regulation, REGULATION (EU) 2016/679, Recital 26)

The most common occurrence of pseudonymised data in scientific studies is that the data for analyses only contains an ID number or similar, and there exists a separate list which matches these ID numbers to names and/or addresses of identifiable persons.

In addition, information which applies only to very few persons makes data identifiable. For example, birth date in combination with sex and postal code is so specific that this make someone easily identifiable.

Any collection, storage and processing of personal data (including pseudonymised data) requires the consent of the respective person.

# Data protection consequences and recommendations

If the purpose of the study allows, collect, store and process only anonymized data.

Of course, even for anonymized data, usually, you need informed consent. This requirement does not result from data protection regulations, but from the rights of participants in scientific studies as detailed in the Helsinki declaration (Oct. 2013 version), particularly numbers 25-32. The informed consent forms have to be kept strictly separate from the other data, and strictly confidential, e.g. in a locked cabinet or similar with defined access opportunity (who has access to it).

If you need personal information, e.g. for a planned follow up study, then store and process only pseudonymised data, using a subject ID and prepare a list (on paper or electronically) which maps subject IDs to personal details, such as name and address (key list or pseudonym list). For this key list the same recommendations are made as for the informed consent forms (see above). If the list is maintained electronically, then the list must be encrypted by an appropriate tool. It must be documented who has access to the list and/or encryption/decryption keys. The pseudonymised data must be stored on a device with is password protected. It should be documented and restricted who has access to this data.

Any pseudonymised data should be anonymised as soon as possible. Study subjects must be informed when this will happen and have to give consent for this.

Anonymisation of data requires the secure deletion of the key list. In addition, parts of the information which would allow identification of persons, alone or in combination, have to be deleted or transformed in a way that make the information really anonymised.

The following table gives an overview about the three types of data which are of relevance

|  |  |  |  |
| --- | --- | --- | --- |
| Characteristic of data processing and storage | Informed Consent | Key List (Pseudonym List) | Scientific Use Data (SUD) |
|  | Signature (on paper) or active selection of agreement (online) | Name (contact data if needed) plus ID | Only ID, no personal details which allow identification; any other data that is needed for the research |
| Pseudonymised data processing and storage | If on paper: keep separate, confidential (locked), restricted access;  electronic with personal details: separate, encrypted, restricted access | Keep separate, conficential (locked or encrypted), restricted access | Usually restricted access.  If de-anonymisation is unlikely, consider open access or restricted access |
| Anonymized data processing and storage (including anonymous data) | If consent is given in online survey, avoid collection of details which allow identification ⇨ consent may be recorded in SUD  Otherwise: as above | Must not exist | as above |
| Storage duration | Usually 10 years (good scientific practice) | Delete/destroy as soon as possible (anonymisation; specify when this will happen) | After anonymisation: unlimited, but at least 10 years (good scientific practice) |

**Data sharing** for scientific use data (SUD): If de-anonymisation of the data is unlikely you could (or even should) consider open access to the dataset or restricted access for which you have to set up rules or regulations of access and processing (for example by contract or general regulations). In any case you should collect the binding assurance of the potential user that he/she will not attempt to de-anonymize the data and he/she will not re-distribute the dataset.

# Data documentation

According to Good Scientific Practice all relevant data has to be kept for 10 years. This applies to both, to the anonymised data and to the separately kept consent forms. A procedure should be established that ensures the secure deletion of the consent forms and any other personal data after 10 years.

# Responsibility

Adherence to data protection rules and good scientific practice is the responsibility of the researcher.

# References

## Links to the GDPR

<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>

<https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679>

## Declaration of Helsinki

Even though the Declaration of Helsinki applies, strictly speaking, to medical research, the principles laid down should also apply to any research involving human subjects as appropriate.

<https://pdf-it.dev.acw.website/please-and-thank-you?url=https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/&pdfName=wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects>

<https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/International/Deklaration_von_Helsinki_2013_20190905.pdf>