Additional instructions for planning the management of sensitive and confidential data 2019

Instructions for references: Working group for drawing up instructions for the management of sensitive data (2019) under the Open Science and Research Initiative. Instructions for planning the management of data containing sensitive and confidential data, Tuuli project.
Zenodo: 10.5281/zenodo.3247282

Introduction	
	These instructions pertain to datasets that include sensitive and confidential research data. Research organisations provide researchers with guidance on complying with the principles of data protection and information security. In addition, other key contact details for organisation-specific support services can be found in each organisation's general instructions pertaining to the matter, also known as data management plans. Several organisations may also have combined these instructions with their data management plan. Please familiarise yourself with the data management guidelines of your organisation.
1. General description of data	
1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Also give a rough estimate of the size of the data produced/collected.	Specify all dataset types that contain personal, sensitive or confidential data. Identifying the sensitive components of research data is particularly important, as the planning of data management focuses on the identification and management of related risks. As regards personal data, specify the party serving as the *controller of the data file*. Sensitive and confidential data is data whose disclosure may cause harm: • Sensitive personal data; no comprehensive listing of sensitive personal data can be drawn up. **The parties conducting the research are responsible for identifying data whose disclosure could harm the study subjects.** ○ Sensitive data may be related to health or the risk of developing a disease, sexual orientation, ethnic background, trade union membership or religious conviction.

	• Sensitive species data, such as data concerning endangered animals and plants, data related to nature conservation or biosafety • Other confidential data, such as patents, data related to national defence, organisational data or trade secrets *Tips* Personal data encompasses all data from which a person can be identified either *directly* or *indirectly*. • *Direct identifiers*: name, phone number, personal identity code, image, audio, fingerprint, dental chart, MRI image • *Indirect identifiers*: gender, age, education, professional status, nationality, location data, career history, system log data, marital status, place of residence, vehicle registration number Links: What is personal data (Finnish Social Science Data Archive), Processing of personal data (Office of the Data Protection Ombudsman)
1.2 How will the consistency and quality of data be controlled?	Consider how data minimisation, pseudonymisation or anonymisation will affect data quality. https://www.fsd.uta.fi/aineistonhallinta/en/anonymisation-and-identifiers.html#when-are-data-anonymous-and-when-pseudonymous https://tietosuoja.fi/en/pseudonymised-and-anonymised-data
2. Ethical and legal compliance	
2.1 What ethical issues are related to your data management, for example, in terms of handling sensitive data, protecting the identity of participants, or gaining consent for data sharing?	Check: *Identity of the controller of the data file* Processing of special categories of personal data • When is the processing of special categories of personal data permitted? Frequently asked questions of the Office of the Data Protection Ombudsman Guidelines for ethical review in human sciences under reform (in Finnish only) Finnish National Board on Research Integrity (TENK) National Committee on Medical Research Ethics TUKIJA Read the guidelines of your own organisation.

2.2 How will data ownership, copyright and intellectual property right (IPR) issues be managed? Are there any copyrights, licences or other restrictions which prevent you from using or sharing the data?	Agreements on data ownership and other intellectual property rights must be concluded before commencing any actual research activities. Template agreements and consultation are provided by your own research organisation.
3. Documentation and metadata	
3.1 How will you document your data in order to make it findable, accessible, interoperable and reusable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?	When describing data, please remember that file and folder names as well as variables and metadata may contain personal or sensitive data. Even if your research data contains personal data, related metadata can be published if it does not contain identifiers which could be used to identify a study subject. • Making a research project understandable - Guide for data documentation • Data Management Guidelines
4. Storage and backup during the research project	
4.1 Where will your data be stored, and how will it be backed up?	According to the General Data Protection Regulation (GDPR) of the EU, risks related to the processing of personal data must be assessed before any such processing. Familiarise yourself with the data protection and risk management guidelines of your organisation and consider the following: • Which freedoms and rights of the data subject can be compromised by the processing? • What kind of damage can the planned processing of personal data cause to the data subject? • What kind of damage can the inappropriate disclosure, destruction or corruption of the data cause to the data subject? • What kind of risks must your data be protected against? • What measures are used to manage identified risks? • What is an accepted level for the probability and impact of residual risks? After completing the assessment, verify with the data protection officer of your organisation whether your data requires an impact assessment in accordance with the GDPR.

Please also find out whether the funder of your research, owner of the data or any other external party has any claims concerning the data.

The risk assessment is used to determine the required protective measures for the entire lifecycle of the data (see also section **4.2**).

Consider the following:

- Which storage services and equipment are used during the research?
- Who is responsible for the maintenance of the storage services used?
- How are backup copies made in the storage services used?
    - Who is responsible for making backup copies?
    - Where are the backup copies stored?
    - How often are backup copies made?
    - How long are backup copies stored?
- Do the storage services maintain a log on data use?
- Can the data be accessed remotely?
    - If so, how is remote access protected?
- Is data encryption necessary?
    - If so, please consider the following:
        - Which parts of the data are encrypted?
        - Which encryption tools are used?
        - Who manages encryption keys and passwords?
- How are the facilities used for processing the data protected?
    - Can the doors of the facilities be locked?
    - Is every holder of access rights known?
    - Does the property have recording video surveillance?
    - Are there burglar-proof storage fixtures/fittings and facilities available for physical material and storage equipment?
    - Are the workstations equipped with screen guards?
- After they are no longer needed, how will data and copies be disposed of in a safe and secure manner?

Please read your organisation’s instructions on storage services and tools with which information security is ensured in the processing of data. Also find out the service address of your organisation’s data protection and IT units.

Further information:

Risk assessment (Office of the Data Protection Ombudsman)

	Impact assessment (Office of the Data Protection Ombudsman)
4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?	Access to personal data must be limited to those individuals for whom it is necessary in order to carry out the research. Please take into account that this group also includes the parties that maintain the services and equipment used and other external service providers, if any. Consider the following: • How, with whom and in what way is it necessary to process the data? o Who can be granted user rights? o What kind of use is permitted? o How do user right requirements vary by data? o Is it necessary to share the data with partners or service providers? o Who can transfer data from one party to another and on what grounds? • How is user and access control implemented? o Who is the person responsible for access rights? o Are user and access rights clear, modifiable and removable? o Are the grounds and restrictions for the rights regularly verified? • How is the use of data and its appropriateness monitored? o How often are user rights to the data checked? o Where and to whom is the data or parts of it copied? o How are data copies managed? o How are data copies erased after the related access right expires? o How are the purposes of use for the data verified? • Are the skills of the individuals processing the data, as well as the instructions pertaining to such processing, up to date in terms of data protection and information security? Please familiarise yourself with the principles, guidelines and tools related to the management of access rights in your organisation. Find out how misuse and damage related to personal data are reported in your organisation.
5. Opening, publishing and archiving the data after the research project	
5.1 What part of the data can be made openly available or published? Where and when will the	Datasets that contain personal data can only be opened after anonymisation. Anonymisation is also recommended due to the fact that anonymised data no longer constitutes personal data. Therefore, it is no longer subject to data protection legislation, which makes accessing and sharing it possible. Pseudonymised data still constitutes

data, or its metadata, be made available?	personal data, which is why it cannot be opened. However, datasets that contain personal data can be disclosed to interested parties who have been granted a permit for a purpose corresponding with the original grounds for processing. The original grounds for processing datasets containing personal data, such as statutory grounds or consent, can restrict any further use. For example, if the original consent form does not refer to the further use of the data, opening the dataset may require requesting fresh consent from the data subjects. Datasets that contain personal data can be opened or published in the following ways: 1. Data anonymisation and opening the anonymised data in a data archive 2. Publishing key metadata (in a research database or another publishing service) and making the actual data available subject to permission from the producer of the data or a reliable data repository *Links* - Pseudonymised and anonymised data - Anonymisation and personal data - Metadata publishing service Etsin
5.2 Where will data with long-term value be archived, and for how long?	The Ministry of Education and Culture offers institutions of higher education and research institutions a service for the long-term preservation of research data (Digital Preservation Service for Research Data (PAS)). Each organisation independently determines their process for identifying research data that will retain its value for a longer period and transferring it to the service. Depending on the guidelines of individual organisations and research permits, datasets containing sensitive personal data can also be stored in the service. Archiving datasets that contain sensitive personal data requires a storage permit from the National Archives of Finland. The data must be minimised before storage. The further processing of such data requires a research permit. Traditionally, the recommendation is to destroy all sensitive data after the research project has ended, as storing it is risky and requires special arrangements. This is why it is important to plan the safe disposal of the data. For example, simply deleting a file and emptying the recycle bin on one’s computer does not mean that the file has been permanently destroyed. Deleted data can be recovered even

	after reformatting the hard disk. There is software available for the permanent disposal of data based on, for example, overwriting or demagnetising hard disks. Storage devices can also be mechanically crushed into an unreadable state. Anonymised data no longer contains personal data and is therefore not subject to data protection legislation. *Tips* • Please remember that the anonymisation and disposal or archiving of data must be carried out by the expiry of the relevant research permit. • Genuine anonymisation requires that both direct and indirect identification are made impossible, in addition to which the identification key must be destroyed. • Many institutions of higher education and research institutions have internal guidelines for the destruction and disposal of storage equipment. *Links* • Disposal of research data
6. Data management responsibilities and resources	
6.1 Who will be responsible for specific tasks of data management during the lifecycle of the research project? Also estimate the resources (e.g., funding, time and effort) required for data management.	Who is **responsible** for the management of sensitive and confidential data as well as monitoring its implementation throughout the lifecycle of the data? • Who is responsible for *data protection* (see section 2) and *information security* (see section 4)? When planning the **resources** needed, the following must be taken into account: • Costs of data minimisation, pseudonymisation and anonymisation, or the time spent and software required for the task • Requirements set by a higher level of security for activities and solutions used, as well as related additional costs