Data Stewards and Data Management Principles
(LTAT.02.014)

Priit Adler
Research Fellow, University of Tartu
Elixir EE, BIIT

2021.03.23
Recap

Data reusability
• ‘plurality’ of rich metadata
• Licence
• Community based standards
To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

https://www.go-fair.org/fair-principles
**FAIR principles**

**To be Findable:**
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

**To be Accessible:**
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

**To be Interoperable:**
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

**To be Reusable:**
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards
Sensitive data
Learning outcomes

• what sensitive data are
• what data protection laws are and how they apply to research data
• what types of research will be subject to ethical review
• what consent documentation is appropriate to ensure you meet your ethical and legal obligations
• the key aspects of data management needed to safeguard sensitive data

https://mantra.edina.ac.uk/protectionrightsandaccess/
When conducting research there are a number of ethical and legal considerations which may affect how your research data are collected, managed and shared.

Understanding your obligations as a researcher will enable you to adopt appropriate data protection-compliant strategies and help you to follow good practice in research integrity and research data management.

https://mantra.edina.ac.uk/protectionrightsandaccess/
We use **sensitive data** as a catch-all term to refer to:

- research data containing information which can be used to identify a human data subject;
- confidential data, including data generated or used under a restrictive commercial research funding agreement;
- ecological data, where the release of data may adversely affect rare or endangered species of plants or animals;
- data likely to harm an individual or community, or have a significant negative public impact, if released.

Other kinds of research data may also be considered sensitive, and researchers should therefore use their own judgement to determine whether research data should be considered as sensitive.

https://mantra.edina.ac.uk/protectionrightsandaccess/
General Data Protection Regulation

applies to the **processing** of **personal data**:

- about people residing inside and outside the European Economic Area (EEA), by people within the EEA
- about people inside the EEA, by people outside the EEA

[https://mantra.edina.ac.uk/protectionrightsandaccess/](https://mantra.edina.ac.uk/protectionrightsandaccess/)
Personal data

information about identifiable living people who can be identified using the data, either directly or indirectly.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Processing

doing anything with a person's information, including:

• collection,
• storage,
• analysis,
• sharing,
• deletion, and
• destruction.
For research, this means that research data containing **direct identifiers** (e.g. a person's name, address, or other unique identifier such as their NHS or Social Security number) will be subject to data protection laws.

Even if research data contain no direct identifiers, an individual may still be identifiable using a combination of **indirect identifiers** (e.g. health, economic, cultural or social characteristics) and so data protection laws will therefore still apply.

There are additional requirements when handling *special category data* relating to a person's racial/ethnic origin, political opinions, religious/philosophical beliefs, trade union membership, genetic and biometric data, physical or mental health, sex life, and sexual orientation.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Data protection principles

- Lawfulness, fairness & transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity & confidentiality (security)

Designed to ensure that people's data are handled with due care and attention. Additionally, you are required to maintain an audit trail and be able to demonstrate your compliance with these principles.

To understand how data protection laws apply to you and your research data, you should consult your own institution's guidance on data protection and GDPR.

For research data relating to people outside the EEA you must also ensure you are aware of and comply with applicable data protection laws in the respective region. For example, the Health Insurance Portability and Accountability Act (HIPAA) in the US.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Research ethics

In addition to any legal obligations, research involving human participants, human material, personal data or live animals will normally be subject to ethical review to ensure the proposed study is conducted ethically.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Research ethics

Ethical principles are in line with or extend beyond legal principles, and cover:

• the purpose and nature of the research itself
• how consent is communicated and recorded
• what data need to be safeguarded during analysis, and destroyed after use

Our local ethics committee: [https://www.ut.ee/et/teadus/eetikakomitee](https://www.ut.ee/et/teadus/eetikakomitee)

[https://mantra.edina.ac.uk/protectionrightsandaccess/](https://mantra.edina.ac.uk/protectionrightsandaccess/)
Informed consent

Informed, voluntary and fair consent to participate in a study is the cornerstone of ethical research involving people. It is intended to ensure that the rights of individual participants are respected and is closely linked with the participant information sheet.

It is through this ethics consent process that research participants can understand what taking part in a specific study will mean for them. Each person can then choose whether to participate using the consent form.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Consent documentation

In order that research participants' consent is informed, voluntary and fair your consent documentation should include:

• a participant information sheet and
• a consent form signed by the participant

https://mantra.edina.ac.uk/protectionrightsandaccess/
The **participant information sheet** is used to inform prospective participants about the study. The information should be written in clear and easy to understand language and should cover the following:

- What the project is about
- What their participation will involve
- Any risks involved for participants and safeguards to minimise those risks
- Assurances about data security and participant confidentiality
- How data will be used in the study (e.g. tables of aggregated data for published articles, reports and presentations)
- Proposed plans for archiving data at the end of the study and potential future secondary re-use of data
- Details of the organisation overseeing the research
- Who to contact for more information about the study

The **consent form** is used to verify that the research participant understands and agrees to participate in the study. The consent form should cover the following points (but not limited to):

- The participant has read and understood the participant information sheet.
- The participant has been given the opportunity to ask questions.
- The participant understands that participation is voluntarily.
- The participant understands that they may withdraw from the study at any time without giving reasons and without penalty (where applicable).
- The participant understands how the data will be managed, shared and archived (as detailed in the information sheet).
- Signature and date of signing for the participant and the researcher.

[https://mantra.edina.ac.uk/protectionrightsandaccess/](https://mantra.edina.ac.uk/protectionrightsandaccess/)
Privacy by design

Data protection law requires you to adopt a ‘data protection by design and default’ approach.


https://mantra.edina.ac.uk/protectionrightsandaccess/
Privacy by design

This means that when planning a research project, you should be thinking about how you will secure your research data and protect the privacy of your research subjects from beginning to end.

For example, think ahead and plan how you will:

• collect and store the data
• manage files and copies
• control access permissions
• prepare data for archiving/sharing at the end of the project

https://mantra.edina.ac.uk/protectionrightsandaccess/
For research involving personal or sensitive data it is recommended for you to complete a **data protection impact assessment** (DPIA).

Conducting a **DPIA** will help you to

- Ensure compliance with applicable legal, regulatory, and policy requirements for privacy;
- Determine the risks and effects; and
- Evaluate protections and alternative processes to mitigate potential privacy risks.

A DPIA should be considered in addition to writing a data management plan (DMP) at the beginning of the project.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Safeguarding sensitive data

When working with data about people, data protection laws require you to apply appropriate "technological and organisational measures" to ensure personal data is handled appropriately and securely.

There are a number of strategies that you can adopt to safeguard the privacy of your research subjects, and these include:

- Data minimisation
- Data retention limits
- Secure data transfer
- Encryption
- Access controls
- Anonymisation

https://mantra.edina.ac.uk/protectionrightsandaccess/
Data minimisation

a key principle in data protection law, relating to the amount of identifying information that you collect and hold about people.

In practice this means you shouldn't collect or keep identifying information longer than is necessary.

Think of the following:

• if personal information isn't needed, don't collect it
• if it is needed, keep it securely
• periodically review whether you are retaining unnecessary identifying information
• when identifying information is no longer needed, safely remove, delete or destroy it

https://mantra.edina.ac.uk/protectionrightsandaccess/
Data retention limits

Deciding how long you will retain identifiable data before removing direct identifiers, applying more complex anonymisation techniques, or deleting the data altogether will help you to comply with data protection laws.

Data retention limits are common practice when working with restricted data from third parties. Access to data shared under a specific license or data sharing agreement commonly requires data to be securely deleted following an agreed period for conducting the research.

When considering the deletion of sensitive data you need to be aware that standard methods for deleting files (e.g. moving files to the recycle bin and emptying it) are not secure, as deleted files may be recovered. It may be necessary to conduct a risk assessment on the data, including assessment of the type of storage used, to determine the most appropriate method for deleting or destroying sensitive data securely.

You should also plan the retention of original collection instruments, such as paper questionnaires or interview recordings. Once these are transferred into an analysis package or a transcript and quality assured or validated, there may no longer be a reason to retain them.

Questions of which data to keep and for how long need to be considered in the context of your ability to maintain the confidentiality of your subjects information, and should be planned in advance.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Secure data transfer

Data protection laws place restrictions on the transfer of personal data, and personal data should not be transferred outside of the EEA.

The UK’s Information Commissioners Office (ICO) provides detailed guidance on data transfer restrictions, including outlining conditions under which personal data may be transferred:

International transfers under GDPR (ICO)

Before considering the transfer of personal data, you should think through whether the transfer of identifiable data is necessary. For example, can data be de-identified or anonymised?

If data cannot be made unidentifiable then you must ensure you have authority to transfer the personal data, and that there are appropriate safeguards in place to protect the data before, during and after transit.

You should make sure that any system you choose for the transfer of sensitive data is secure. Where possible, it is recommended to use contracted services which can certify they are GDPR compliant.

Widely available services (e.g. Dropbox, Google Drive, etc) provide no user control over where or how data will be processed, and use of such services is therefore not recommended without an organisational contract.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Encryption

Encryption provides protection by ensuring that only someone with the relevant encryption key (password) will be able to access the contents.

Encrypting sensitive data will ensure that it cannot be accessed by non-authorised people. Encryption allows you to secure sensitive data by:

• encrypting a single file
• creating an encrypted container (i.e. a folder) on a hard drive
• encrypting a computer / laptop
• encrypting other devices (e.g. smartphones, tablets)

Note, encrypting a single file is not recommended. Rather, it is advised to create an encrypted container and store sensitive files securely inside it.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Access controls

You should consider applying access controls to ensure that only authorised individuals are able to access sensitive data. For example, standard access control methods include the following:

- password protection
- management of shared folder permissions
- Special software solutions

https://mantra.edina.ac.uk/protectionrightsandaccess/
Anonymisation

Anonymization is a process by which identifying information in a dataset is removed or masked, and is used primarily to enable data to be shared or published without revealing the confidential information it contains, while limiting the loss of information.

Where possible, **direct identifiers** (e.g. names, addresses, telephone numbers, account numbers, etc.) should be removed as soon as the identifying information is no longer needed, by deleting them or replacing them with pseudonyms. For qualitative data, replace or generalise identifying characteristics when transcribing interviews.

De-identified data that can be re-identified using a linkage file (i.e. information linking data subjects to identifiable individuals) is known as pseudonymised data. **NOTE: In this instance, the linkage file should be encrypted and stored securely and separately from the de-identified research data.**

Identification of individuals in pseudonymised or de-identified data may still be possible using combinations of **indirect identifiers** (e.g. age, education, employment, geographic area, medical conditions, etc). Further, data and outputs (e.g. tables of results) containing small cell counts may be potentially disclosive, particularly where samples are drawn from small populations or include cases with extreme values or relatively rare characteristics.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Direct or indirect identifier?

Which of the following are direct identifiers? Please select all that apply.

- [ ] medical condition
- [ ] NHS or social security number
- [ ] job title
- [ ] personal address
- [ ] ethnic origin
- [ ] name

https://mantra.edina.ac.uk/protectionrightsandaccess/
Direct or indirect identifier? (answer)

Which of the following are direct identifiers? Please select all that apply.

- [ ] medical condition
- [x] NHS or social security number
- [ ] job title
- [ ] personal address
- [ ] ethnic origin
- [x] name

https://mantra.edina.ac.uk/protectionrightsandaccess/
Making data available to others

In terms of data sharing, there is currently a tension between the risks of personal disclosure and the pressure to make data openly available as part of the record of research.

If you plan to share or publish your data you must ensure that your data are appropriate and safe to share. For example, you should consider whether the data can be adequately anonymised, and whether anonymised data will remain useful.

After applying methods to de-identify and anonymise sensitive data, there may still be a risk of re-identification. You should then consider applying access controls to ensure the data are shared appropriately and securely. This may involve finding a data repository which can provide suitable access controls.

https://mantra.edina.ac.uk/protectionrightsandaccess/
Controlled access
Registration required
Open & freely available

https://www.youtube.com/watch?v=gFNzhqrPIs&t=10s&ab_channel=LibraryResearchSupport
Data sharing agreements

When sensitive data need to be shared, for example between one organisation and another organisation or third party, a data sharing agreement may be used to specify the conditions under which the data are to be used.

“[A data sharing agreement] sets out the purpose of the data sharing, covers what is to happen to the data at each stage, sets standards and helps all the parties to be clear about their respective roles. It helps you to demonstrate your accountability under the GDPR.”
ICO Data sharing code of practice

https://mantra.edina.ac.uk/protectionrightsandaccess/
What is the purpose of the processing?*

Choose from the dropdown menu the scenario relevant for the data sharing

Choose an option

Please provide a short description of the aim of the processing that can serve as the preamble of the contract (max. 500 characters)*
RDMkit – Data protection
How do I protect research data under the GDPR?

Where scientific research involves the processing of data concerning people in the EU, it is subject to the General Data Protection Regulation (GDPR). The GDPR applies a “special regime” to research, providing derogations from some obligations given appropriate criteria are met and safeguards are in place. The criteria is to follow standards in research method and ethics, as well as to aim societal benefit rather than serving private interests in research. The safeguards are a multitude and include:

- data collection with informed consent under ethical oversight and accountability,
- ensuring lawful processing and exchange of human-subject information,
- putting in place organisational and technical data protection measures such as encryption and pseudonymisation.

The practical impact of the GDPR on research is, then, establishing these safeguards within projects.
Considerations

Seek expert help for the interpretation of GDPR legal requirements to practicable measures.

• Research institutes appoint Data Protection Officers (DPO). Before starting a project you should contact your DPO to be informed of GDPR compliance requirements for your institution.

• Each EU country has its own national implementation of the GDPR. If your project involves a multi-national consortium, the requirements of all participating countries need to be met and you should inform the project coordinator of any country-specific requirements.

• Legal offices in research institutes provide model agreements, which cater for various research scenarios and consortia setups. You should inform your local legal office of your project’s setup and identify the necessary agreements to be signed.
Can we outline 5 steps to follow so to be compliant with the new GDPR?

1. First of all – don’t panic! The foundation of data protection rules is still the same as before.
2. If the processing of personal data is a large-scale process in your company or involves considerable risks, we suggest to conduct an integrated assessment of your data processing.
3. Observe your entire work order, information systems, and document blanks from the perspective of the new data protection rule.
4. Be sure to check the data portability in your work processes and information systems. I forecast that this will be the costliest and most time consuming of the implementing activities.
5. Regardless of whether or not the GDPR requires you to determine a data protection officer, find a specialist to assist you. Send your own employee to a trusted training course or make sure an external consultant has undergone relevant training.

https://www.aki.ee/et
Choice of local references

• https://courses.cs.ut.ee/2020/privacy_tech/

• https://www.aki.ee/et

• https://minu.etais.ee/policy/privacy/
• https://www.ut.ee/et/andmekaitsetingimused
Homework

Prepare 20 - 30 min presentation for the next lecture:

• RedCap database
• Amnesia tool