

How to write a Research Data Management Plan

- Compiled (based on NCN guidelines)
Agnieszka Wasilewska



How will new data be collected or produced and/or how will existing data be reused?

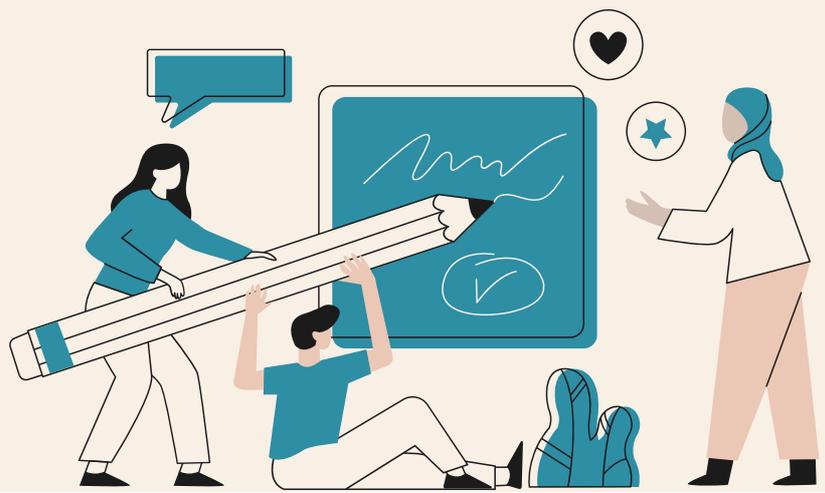
Describe the data you will collect or create, including content, coverage and type of data. Also briefly describe the type of data, such as text, images, photos, measurements, statistics, physical samples or codes.

New data - a general description of how this data will be obtained, how it will be extracted, a description of the methodology, equipment used, processes employed.

- Consider how your data can complement and integrate with existing data, or if there are any existing data or methods you can reuse.
- Indicate which data has long-term value and should be shared and/or protected.

Reuse of existing data - indicate the source of the data, the conditions and any limitations on its reuse.

- Demonstrate that the data comes from reliable sources.
- Will the data owner transfer reuse rights?
- Data from archives, libraries - are there no restrictions on its reuse, under what conditions can it be used, do you need to have permission to collect and reuse the data?



What data (for example the kinds, formats, and volumes) will be collected or produced?

Volume:

- Estimated data size in MB / GB / TB.
- Indicate the proportion of raw data, processed data, and other secondary outputs (e.g., reports).
- Consider the implications of the volume of data in terms of storage, access and protection. Do you need to include additional costs?

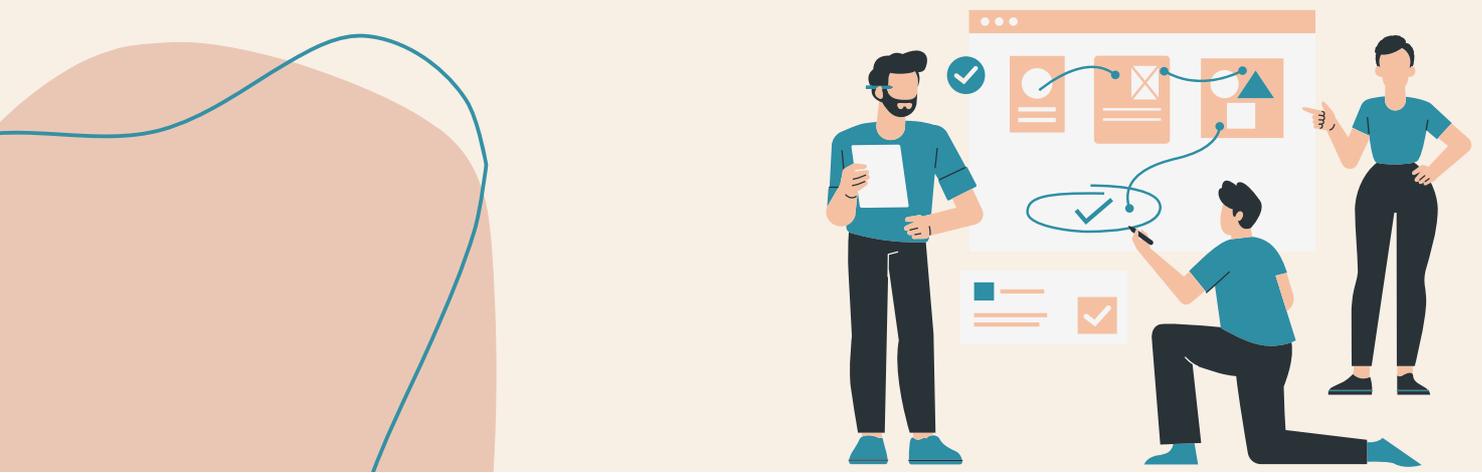
Format:

- What format the data will be in - use open and standard formats.
- Tables: it is recommended to use files with formats: CSV, TSV, SPSS portable. Files are not recommended: XLS, XLSX.
- Texts: it is recommended to use files in formats: HTML, RTF, PDF. Files are not recommended: DOC, DOCX.
- Media: it is recommended to use files in formats: MP4, Flacc. Not recommended files: QTFF.
- Image: it is recommended to use files with formats: TIFF, JPGE2000, PNG. Not recommended files: GIF, JPG.
- Structured data: it is recommended to use files with formats: XML, RDF. Not recommended files: RDBMS.



What metadata and documentation (for example methodology or data collection and way of organising data) will accompany data?

- **Metadata standards** - this will make the data findable not only by humans but also by search algorithms. Research data repositories have the appropriate formats for creating metadata. Data placed in a recognized, domain-specific research data repository will have a metadata description that conforms to the standard.
- **Documentation** - describe the types of documentation that will accompany the data to help users understand and reuse it. Documentation should describe the general context of the data collected - information on the source of the data, information on methodology, on procedures, definitions of variables or units of measurement.
- **Folder and file organization** - use of a consistent consistent structure, naming of files to reflect content, standard date format - so that content can be easily found and sorted.
- **Naming convention** - use of controlled vocabularies used in the discipline.



What data quality control measures will be used?

Data quality - subjective and objective usefulness, accuracy, completeness, consistency, timeliness, uniqueness

New data:

- Procedures used - type of data production/collection control measures and their frequency.
- Describe how data quality and consistency are planned to be controlled and documented, e.g., equipment calibration processes, multiple measurements, standardized methods with forms, questionnaire templates, screening questions.

Reused data:

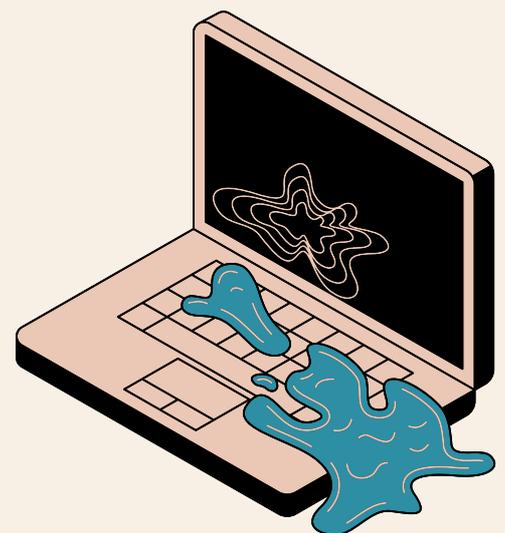
- The source should be described.
- Data should come from a reliable source.



How will data and metadata be stored and backed up during the research process?

Where will data and backups be stored during the study?

- Carefully choose the hardware on which data will be collected and stored. This includes both servers and storage space (storing data on private computers, laptops, portable drives, USB devices is not recommended).
- Cloud storage - with the proviso that the servers will be in Europe. The use of private commercial clouds is not recommended. Ideally, the service should be governed by a bilateral agreement with the university (i.e. OneDrive).
- TASK IT Center - offers data storage for NCN projects carried out by UG employees - 100 GB + 200 GB back-up.
- Backups - it is recommended to follow the 3-2-1 rule - at least 3 copies, on two different media, one of the media should be in a completely different location.
- How often the backup will be done, whether it will be an automatic/manually controlled process, where and by whom, how often the password will be changed..

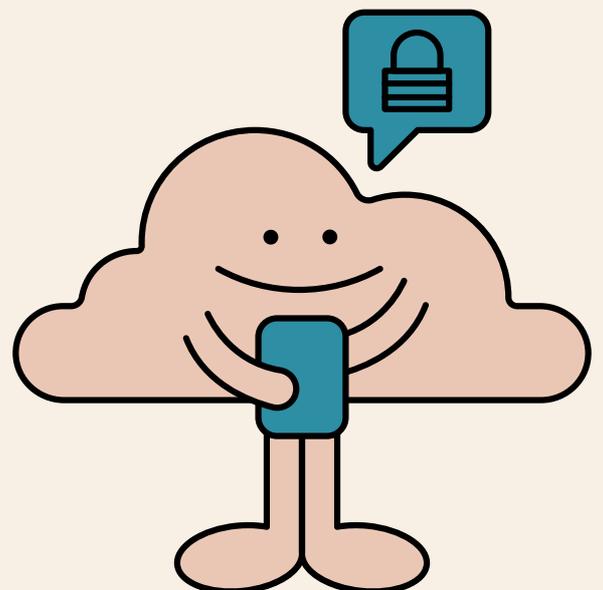


How will data security and protection of sensitive data be taken care of during the research?

- Who will have access to the data during the project, how will access be controlled?
- Anticipate possible threats and risks and how to deal with them.
- If the data is confidential/sensitive, outline any appropriate security measures and note any formal standards you will follow, e.g., encryption, anonymization, tightly controlled access.

Physical security - against unauthorized access, destruction, media damage, theft.

Digital security - back-up, strong password, antivirus, firewall.



If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

If personal data were not obtained, we mark N/A (no personal data will be processed).

We refer to the institution's Data Protection Policy:

- UG Data Protection Officer.
- Data will be collected and processed in accordance with the Data Protection Regulation: UG Rector's Order No. 163/R/21 on the Personal Data Protection Policy at the University of Gdansk.
- Personal Data Protection Policy at the University of Gdansk.

- Was informed consent obtained from study participants for participation in the study and data processing?

- Security - anonymization, pseudonymization, encryption.



How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

- Identify the owners of rights to any data acquired and produced.
- Collaboration with another unit or use of another unit's equipment - the sharing of intellectual property rights should be agreed with the collaborating unit.
- On what terms the data will be shared - it is recommended to use standard Creative Commons licenses.
- Whether there are legal restrictions on the reuse of existing data - consider copyright. If using data from libraries or archives - adhere to rules and regulations.
- Ethical aspect - whether there are ethical contraindications, whether an ethics committee review is needed, approval from the Bioethics Committee.
- Consents and permits for research - whether the research requires permits from specialized bodies.



How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Where will the data be shared - how will potential users learn about the data?

- Recommended - research data repository.
- The repository should provide a DOI or other permanent unique identifier and comply with the FAIR principle.

When will you make the data available? - Is there an embargo (how long).

- Data should be made available as soon as possible - when the research results are published.
- Embargo - the reason should be justified.

If any collections cannot be made available or access must be restricted, the DMP should explain why this is necessary (e.g., restrictions of a legal nature).

Data should be "as open as possible, as closed as necessary".

- Restrictions on release due to legal reasons, potential for commercialization, confidential, sensitive data, company secrets, etc.



How will data for preservation be selected, and where will data be preserved longterm (for example a data repository or archive)?

Selection:

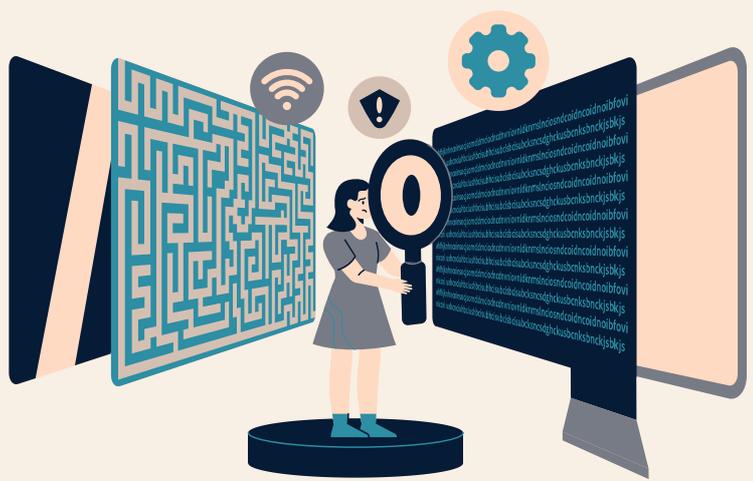
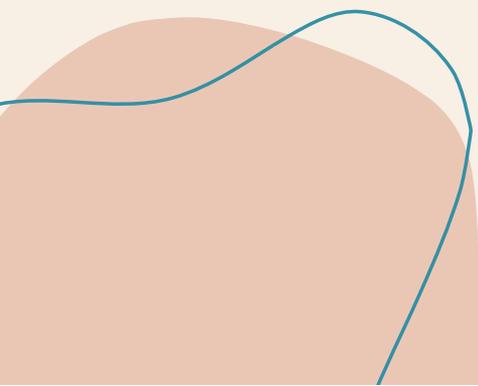
Which data should be retained and which should be destroyed? On what basis do you decide what data to keep, e.g., if the data can be reused to confirm research results, conduct new research. Do you want to share data from an experiment that failed so that others don't duplicate mistakes? Decide which data to keep and for how long.

Which data will be shared?

- Not all data must be made available in the repository - only those selected, cleaned, of the best quality.
- The requirement for sharing applies to data that form the basis of publications, used to verify research results.
- Will raw data also be made available? - If not then the original raw data should be archived.

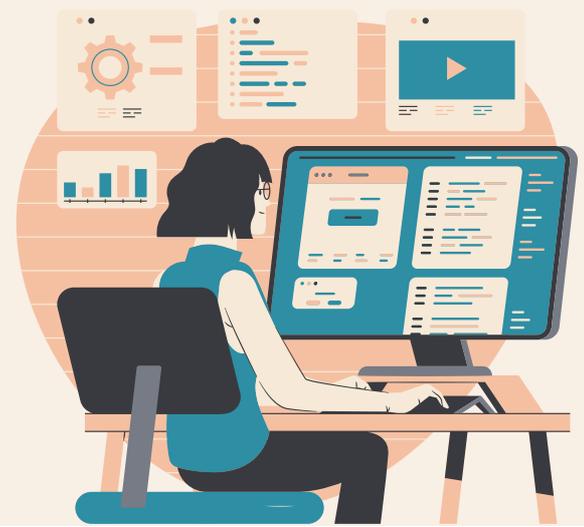
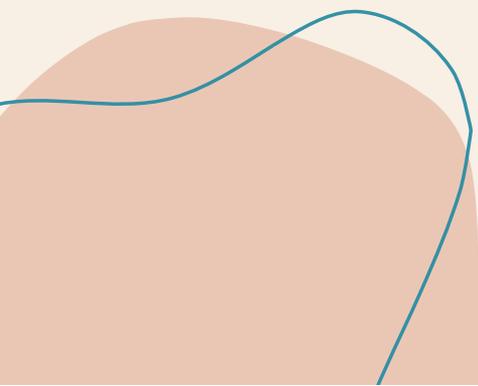
Archiving:

- How long will they be kept - in the sense of NCN, a reasonable retention period is a minimum of 10 years.



What methods or software tools will be needed to access and use the data?

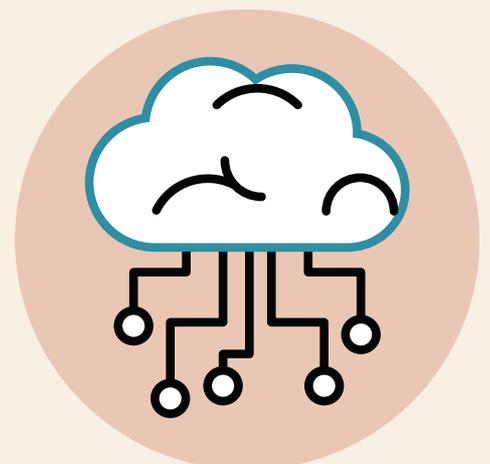
- Will potential data users need specific tools to access and reuse the data?
- Will additional hardware or software be required for scanning or conversion?
- Describe the software used during data collection, tools (e.g. MatLab).
- FAIR principle (interoperability) - is it possible to be used by other people and entities (including machines). If the use of the collection requires the development of a translator, a way of mapping or conversion of formats, this indicates its limited interoperability. Thus, we can talk about full interoperability when the collection we have deposited is available through an API, and its metadata additionally through the OAI-PMH protocol.



How will the application of a unique and persistent identifier (for example: such as a Digital Object Identifier (DOI)) to each data set be ensured?

The identifier is assigned by the repository - when choosing a research data repository, it is important to note whether the repository assigns a DOI or other unique, persistent identifier.

- Persistent identifier - allows you to reach the collection even when the url changes, allows you to locate the data accurately and efficiently, serves as a reference, and allows easy tracking of citations and reuse.



Who will be responsible for data management (i.e. data steward)?

- Who will be responsible for each data management activity?
- How will responsibilities be distributed among partner parties in joint research projects?
- Outline the roles and responsibilities for each activity, e.g., data capture, metadata creation, data quality, storage and backup, archiving and data sharing. Consider who will be responsible for ensuring that appropriate policies are followed.



What resources will be dedicated to data management and ensuring that data will be FAIR ?

Resources needed to implement the plan. These may include storage costs, hardware, software, staff time, data preparation costs for deposit and repository fees (mostly repositories are free). Where dedicated resources are needed, outline them and justify the need for them.

- What resources do we have, and what resources do we still need to acquire?
- What will be the costs associated with ensuring FAIR standards in the project? (Costs associated with long-term storage and sharing of data).



Research Data Management and Open Science Section

About us:

- We provide consultation in research data management.
- We offer assistance in developing DMPs.
- We advise on data security.
- We organize individual and group training courses.
- We manage the University Research Journals platform.

Team:

Agnieszka Wasilewska

Zbigniew Ruszczyk

Agnieszka Kranich-Lamczyk

Katarzyna Cherbetko-Zygmuntowicz

Get in touch with the Data Steward

Main Library, room 0.44

tel. 58 523-32-67

e-mail: research.data@ug.edu.pl

