Ahus - Data management plan for digital research data

Doc . admin.: Subaitha Navaruban Valid from: 19/04/2023 ID: 35413
Approved by: Øystein Mæland Revision deadline: 18/04/2025 Revision: 1.2

Data management plan template for digital research data

A data management plan is a tool for handling research data. The plan describes how data will be handled during the project period and after the project has ended. The purpose is to assess various aspects of the handling of research data, from collection/generation, processing, analyses, documentation, to storage and future sharing of data. A data handling plan should help ensure that research data can be handled legally, structured and securely, and can be stored, reused and understood in the future. The data management plan must be a living document that is regularly updated throughout the life of the project. Projects must deliver the first version of the plan when revising the application. An updated and final version is delivered together with the project's final report. Data handling plans should, as far as possible, be public and published openly so that professional communities can better follow the practices of their colleagues.

Some projects require a data management plan:

For all projects that receive funding from the Research Council, the institution responsible for the project must assess the need for - and the quality of - a possible data management plan. If the project manager decides that the project will not prepare a data management plan, the Research Council must receive a justification for this when the project manager submits a revised application.

EU projects funded by H2020 are required to develop a data management plan within 6 months of receiving funding.

In clinical drug trials, a data management plan must be established.

		You should answer this
1	Responsibilities and administrative information	The data handling plan should contain information about the institution responsible for the project, project manager, project number, project title, financier and version.
2	Collection and/or use of existing data	• Which methods or software are used if new data is collected or produced? • How should data provenance (origin) be documented? • Are there any restrictions related to file format, licensed software or the like for (re)using existing data? • Are there any reasons why existing data sources are not reused?
3	File types and format	• What types of data will the project collect, produce and/or reuse? Examples of types of data can be numerical, text, image, sound, video etc. • What data format is data stored in during collection and analysis, for example . pdf , . xls (x), . doc (x), . txt , . rdf . • On what basis are certain data formats chosen? For example, this may depend on expertise in certain formats, preference for open file formats, standard formats accepted by data repositories , extensive use of certain formats in research environments or formats provided by the equipment or software used.

		• What is the (estimated) volume of data that must be stored during collection and analysis, archived and possibly long-term storage? This can be stated as storage space (bytes), number of objects, files, rows and columns.
4	Documentation, metadata and data quality	Will the structured data be self-explanatory (based on the variable names, codes and abbreviations used)? Are there relevant metadata standards that can be used to describe data quality? Make sure to use metadata at both project level, data set level and variable level that, if possible, follows international standards.
5	Organization of data	• What metadata will be used to help others identify and discover the data? • What metadata standards will be used? Examples can be DDI, TEI, MARC, CMDI. • How will data be organized throughout the project? This could be version control, file structure, file naming conventions, etc. • Is there other documentation that is necessary to facilitate reuse? It can be a description of methodology, information about analysis and protocols, definitions of variables, electronic lab books, code books, readme.txt files etc. • How is the reliability and quality of the data checked and documented? This may include processes such as calibration of measuring instruments, repeated measurements and samples, standardized data capture, validation of data recording, peer review of data or use of controlled vocabulary.
6	Storage and data security during the project	• Where will (meta)data be stored and backed up throughout the project, and how often will this be done? Storing data on laptops , typical external hard drives, USB sticks or the like is not recommended due to less protection and a greater risk of data being lost. • How should data be restored in the event of an accidental event? • Who will have access to the data during the project and how will access be controlled? This is particularly important where the project is a collaboration with several research environments/institutions. • If applicable, how are data security and risk management related to sensitive data, such as personal data and data underlying trade secrets, handled? • Which institutional guidelines for data security are applicable?
7	Rights and sharing restrictions: legal requirements and ethical guidelines	• Which legal entity has rights and/or access to the research data? • Will the data be openly available or with access restrictions, if so what access restrictions? An example is that access to data is only granted via a service with authentication.

		• Will there be any purpose restrictions, for example that the data can only be used for non-commercial purposes, and if so why? • What release statements or licenses should be applied to the research data? • Where the project has several collaboration partners and/or several legal or natural persons with rights to research data; how should rights to control data access be handled in the project? • Where the research data falls under copyright or database protection in the Copyright Act ; which rights apply and how will this be handled in the project? • When using data from a third party; what access and purpose restrictions, if any, apply to this data? • What ethical issues can affect how data is stored and transferred, who has data access to see or use the data, and how long it must be kept? • Which institutional, national and/or international research ethics guidelines apply to the project? Examples could be approval from regional committees for medical and healthcare research ethics (REK) or the Norwegian Food Safety Authority. These points should be described if you handle personal data: • How are the GDPR (Personal Data Protection Regulation) and the Personal Data Act complied with when handling/processing personal data? o Is informed consent used for long-term storage and possible sharing of personal data? o Is anonymisation, pseudonymisation or encryption of personal data being considered for long-term storage and/or sharing? • Should an administered procedure be used for authorized access to personal data?
8	Data sharing and reuse	Remember that data sharing does not mean that you must necessarily share openly, but according to the principles "as open as possible, as closed as necessary" and FAIR (eng: findable , accessible , interoperable , reusable) . Data sharing and reuse must be seen in connection with the other topics in the guide, but is highlighted here as a separate topic because it is an important aspect of the life cycle of your research data. Therefore, you will recognize some of the points that should be described from other parts of the guide. These points should be described: • How should the data be findable and how should it be shared? Examples could be that they are made available in a certified data warehouse (eng: repository), are indexed in a directory,

		that one uses a secure data service, direct handling of data requests, etc. • When will the data be shared? For example, will use be made of an exclusive right of disposal granted by legal legislation that affects the timing of sharing and, if so, why and for how long? Examples could be that you wait until a scientific publication is available or that you want to protect intellectual property rights, such as patent rights, until you have applied for a patent. • Who can reuse the data? If it is necessary to restrict access, for example that only certain groups/environments have access or an agreement on data sharing should be used, it should be explained how and why and what measures are taken to minimize restrictions. • How can the data be reused in another context? For example, is there potential for commercial exploitation? • Do potential users need specific tools, such as software, to access and (re)use the data? The sustainability of the software for future access to the data should be taken into account. • Will a persistent identifier (DOI) for the datasets be used? Persistent identifiers should be applied to metadata and datasets so that they can be retrieved and referenced in a reliable and efficient manner. Use of DOI also ensures that citations and reuse can be tracked. A certified data warehouse for long-term preservation (eng: respository) will often provide this for (meta) data that is deposited there.
9	Long-term preservation	• Which data must be kept or deleted based on agreements, legal legislation and/or guidelines? • Which data must be kept long-term and which criteria are used to select these? • What are the potential future research purpose(s) and/or users of the data? • Where will the data be stored long-term (for example which data warehouse, eng: repository)? If a specific data warehouse is not proposed, the plan should show that the data can be curated appropriately after the project's lifetime. It is recommended to refer to guidelines and procedures at data warehouses , including metadata standards and costs involved.
10	Data management – responsibilities and resources	• Which roles have which responsibility for data handling activities in the project? Examples of activities are data capture, production of metadata, data quality, storage and backup, long-term preservation and data sharing. Responsible individuals should be provided, if possible. • For collaborative projects; How is responsibility for data handling coordinated between partners?

		• Who is responsible for implementing the data management plan and for the plan to be reviewed and regularly updated? In our guidelines, it is the responsible institution that must approve the plan. • How are necessary resources budgeted and covered in the project to prepare data for sharing and long-term preservation (curation)? These can be costs related to storage, hardware, staff time, costs for preparing data for deposit and costs related to preservation at a data warehouse (eng: repository).

For more information and guidance on the requirements for a data handling plan, see the websites of the Research Council , or the websites of NorCRIN .