University of Helsinki Research Data Policy

Contents

1. Introduction

Research data are central to science and research. The production of data requires long-term efforts as well as technical and financial resources. In fact, the reusability of research data has become an increasingly important question in terms both of science and research, and of the impact of research.

For the University of Helsinki as well as its units and researchers, research data constitute a strategic resource and an international competitive factor. The goal of the University is to promote responsible research data management, which is of crucial significance for the findability, accessibility and reuse of research-based knowledge (*for the definition of 'research data' and 'research data management', see the glossary*).

The principles of open science and open research data play a prominent role in the strategic plan of the University of Helsinki for 2021–2030. The University's strategic goals include open research infrastructures and open research data, the implementation of the FAIR Principles in research data management, and the advancement of competencies in the analysis of large and open datasets.

A key goal of the research data policy is to guide everyone involved in research data management to familiarise themselves with current data management requirements as well as to adopt good and responsible practices as part of everyday research activities. Another goal is to make research data management increasingly clear for individual researchers.

From an organisational perspective, the research data policy describes the goals that guide the development of research data services at the University of Helsinki. The goal is for University of Helsinki researchers to have at their disposal infrastructures and services that enable responsible research data management, developed in an economically sustainable manner while taking researchers' needs into account.

In terms of its fundamental goals, the new research data policy of the University of Helsinki does not significantly differ from the previous policy published in 2015. The new policy has been amended to comply with current legislation as well as national and international guidelines and recommendations. In contrast to the previous policy, whose more limited scope only encompassed digital data, the new research data policy applies to all research data and related management principles. Another important change is the further clarification and highlighting of responsibilities.

1.1 Scope of the research data policy

This research data policy covers all research conducted at the University of Helsinki, as well as the research data collected and produced in conjunction with it. This denotes both digital data as well as physical and analogue data, which are all referred to as *research data* in the research data policy (*for the definition of 'research data', see the glossary*).

The research data policy applies to all members of the University community involved in research, including University employees and students as well as those who conduct research on behalf of the University.

1.2 Other policies pertaining to research data management at the University of Helsinki

The principles and goals presented in the research data policy supplement other University of Helsinki policies relevant to the management of research data, including the principles of open science, the data protection principles, the information security policy and the guidelines of the Finnish National Board on Research Integrity for the responsible conduct of research. The research data policy is also in line with the University's data management practices and principles.

Policies related to research data management at the University of Helsinki are guided primarily by EU and Finnish law. By employing policies and guidelines, the University strives to ensure the legality of research activities. In international research and other cooperation, research data management can also be regulated by legislation outside the European Union.

External research funders and partners may also have specific requirements for research data management.

The general objectives for the processing of data stored in University of Helsinki research infrastructures or those on the national or international level are described in the University of Helsinki Research Infrastructure Programme (*for the definition of 'research infrastructure', see the glossary*). In addition, national and international research infrastructures may have their own jointly agreed policies for the processing of the data that they produce. When research infrastructures draw up or update their data management principles, their alignment with the University's research data policy must be ensured.

1.3 Research agreements and agreeing on rights

Research agreements

Concluding agreements and undertakings is an important part of responsible data management, regardless of the source of research funding. Agreements are always needed when research is conducted in a collaboration or other partnership with external parties, such as other universities, businesses, research organisations and hospital districts. Agreements help to safeguard the interests of researchers and the University of Helsinki, manage risks related to research and ensure the legality of activities.

Research agreements must be concluded as early as possible, preferably before the collection or use of research data commences. Agreements are used to ensure that the research data collaboratively collected by research groups is available to all of the participating researchers. Agreements safeguard the continuity of research and the further use of research data, for example, after the conclusion of research projects.

Agreeing on rights related to research data

Many international and national research funders require that the research data and findings that they have funded are open access. Obligations imposed by funders on researchers and the University make it necessary to transfer rights to research data to the extent necessary to fulfil the funding terms. In addition, the sharing of rights is needed, for example, to enable the opening, further use and archiving of research data. As a rule, researchers retain a concurrent right that enables them to continue using the research data.

The rights related to research data and their sharing should be agreed when researchers' employment at the University begins.

Agreeing on the rights related to research data is in line with legislation and the responsible conduct of research. Instead of altering researchers' responsibilities in the management of their research datasets, such agreements are part of it. Sharing rights does not alter researchers' right to be referred to as the collectors or producers of the research data in question.

1.4 Updating and monitoring the research data policy

This document supersedes the current University of Helsinki research data policy that was approved in 2015. An implementation plan for the updated data policy will be drawn up in 2022.

The implementation of the research data policy will be monitored by carrying out assessments at regular intervals, the first in 2023.

The research data policy will be updated by 2025.

2. Principles and goals for research data management

2.1 General principles

Below, the principles and goals for responsible research data management are described in relation to the lifespan of research projects, beginning from the planning of data management and ending in the long-term preservation or destruction of data. The potential for the further use of data must be taken into consideration during research projects.

In general, the responsible management of research data is guided by the FAIR Principles, according to which research data must be *findable*, *accessible*, *interoperable* and *reusable* (*for the definition of the 'FAIR Principles', see the glossary*). Another guiding principle pertains to the openness of research data: as open as possible, as closed as necessary.

2.2 Planning and preparation of data management

- A data management plan is to be drawn up for all research projects. The plan is to be updated during the project whenever substantial changes are made to the content (*for the definition of 'data management plan', see the glossary*).

- Questions related to legislation and research ethics are taken into consideration in the data management plan.
- When processing personal data, the University of Helsinki data protection guidelines on informing data subjects and ensuring the necessary safety measures are to be taken into account. When necessary, a data protection impact assessment must be carried out (*for the definition of 'personal data', see the glossary*).
- The rights and responsibilities related to research are to be agreed within research groups, while the relevant agreements are to be concluded with external parties prior to commencing the collection or use of research data.
- Preparations must be made for the protection and, when necessary, anonymisation of sensitive data as well as the destruction of confidential data, with sufficient resources allocated at the planning stage of research projects (*for the definition of 'sensitive and confidential data', see the glossary*).
- Agreements, consents, undertakings and other documents related to research must be archived so that they are always available to researchers and support services.
- Research infrastructures extensively involved in service provision in particular should draw up a data management policy describing the principles of processing, storing, distributing, preserving and destroying data produced by the infrastructure, as well as the questions of responsibility and ownership related to the processing of data.

2.3 Documentation and processing of research data

- The documentation and metadata associated with research data should follow discipline-specific standards to enable the reuse and further enrichment of the data in future research projects (*for the definition of 'metadata', see the glossary*).
- Appropriate information security must be taken into consideration in the storage and processing of research data.
- The University provides researchers with basic services in the storage and processing of research data. If the quantity of data is high or special computing power is needed, the costs of storage and processing must be taken into consideration at the planning stage.

2.4 Publication and accessibility of research data

- Research data produced at the University of Helsinki and linked to published research results are, as a rule, open and available for shared use. The principle of 'as open as possible, as closed as necessary' is observed when making research data openly available.
- Research data are published in data archives that safeguard the findability of data and enable references to them (*for the definition of 'data archive', see the glossary*).

- Metadata associated with research data must be published whenever possible, either in national or international metadata services.
- Identifiers and licences enabling the further use of data should always be available for open research data (*for the definition of 'persistent identifier', see the glossary*).
- If possible, sensitive data are also made available to other researchers, taking into consideration legal, ethical and contractual restrictions. In this, services that enable safe storage and restricted access with a research permit are to be employed.
- The University has practices and services for data collection and monitoring pertaining to previously collected, produced and opened research data.

2.5 Commercialisation of research data

- In the case of commercially valuable data, measures must be taken to ensure that their use does not endanger commercialisation. For example, research results must be published in compliance with terms and conditions set on confidentiality by the funder as well as the University's guidelines for commercialisation.
- In the case of research that generates commercially valuable research data, the University provides researchers with support and a streamlined service process.

2.6. Destruction or long-term preservation of research data

- The part of research data that is valuable in the long term and that will be preserved, as well as the part that will be destroyed at the end must be specified during research projects.
- Research data that have become useless will be destroyed after the required storage period. Research data that contain sensitive data will be destroyed with particular care.
- The University has a process for the curation and transfer to long-term digital preservation of valuable research data (*for the definition of 'long-term preservation', see the glossary*).

3. Responsibilities related to research data management

3.1 Researchers' responsibilities

- Familiarising themselves with guidelines related to responsible research data management and complying with them
- Planning and implementing data management for their research
- Ensuring that agreements, undertakings, and consents required for research are drawn up and concluded
- Familiarising the members of their research group with responsible data management

when serving as principal investigators

- Ensuring that data that have been agreed to be shared by the group or the collaborative research project are accessible to others
- Conveying the principles and good practices of responsible data management when serving as thesis supervisors
- Updating their skills in research data management on a regular basis

3.2 Responsibilities of faculties and independent institutes involved in research

- Maintaining an overview of research data and their management in the unit, including the unit's agreements and obligations
- Taking the research data policy into consideration in the planning of operations and finances, as well as allocating the necessary resources
- Integrating University-level guidelines and policies into unit operations, taking research field–specific differences into account
- Ensuring that the academic staff and students are familiar with research data management as part of the responsible conduct of research
- Offering and allocating resources for on-site support for research data management together with research support services
- Carrying out preventive risk management and anticipating potential information security incidents (*for the definition of 'risk management' and 'information security incident', see the glossary*).
- Ensuring the implementation of responsible research data management in research infrastructures

3.3 The University's responsibilities

- Engendering the necessary preconditions for the implementation of responsible research data management at the University
- Ensuring that the University-level research data infrastructure is fit for purpose, up to date and sufficiently funded so that services are available at all stages of research projects' lifespans (*For the definition of 'research data infrastructure', see the glossary.*)
- Establishing, in collaboration with academic units, assessment practices and incentives to consider researchers' efforts to promote the sharing and further use of research data as well as skills in research data management as academic merits
- Establishing incentives for University units for responsible research data management
- Promoting opportunities for researchers and support services specialists to specialise in the management of research data by consolidating related specialist roles and developing career paths
- Coordinating development efforts related to research data management
- Drawing up University-level policies for research data management and processes that support their implementation

- Drawing up the necessary guidelines for research data management
- Carrying out and supporting preventive risk management and anticipating potential information security incidents
- Offering researchers and other staff training and orientation related to research data management
- Offering researchers and academic units support in research data management

4. Glossary

Data archive (data repository)

A virtual, typically discipline-specific archive or database where researchers can transfer their research data for sharing, reporting and reuse. Data repositories store research data, make it available and organise it in a logical manner. Data repositories also make it easier to cite research data through the use of persistent identifiers.

Data management plan

A document that describes the research data that will be acquired or produced in the research project, commonly abbreviated as DMP. In connection with data management plans, the term ‘data’ is understood in broad terms, meaning that it encompasses all of the data and resources on which the research results are based. The plan also encompasses codes, software and other methodological descriptions.

In addition, the plan describes how rights related to research data are managed, which agreements are needed, how data protection is ensured, how research data are stored, how research data are opened, or how their findability and use for the verification of the research results and further research is otherwise enabled. While the data management plan is drawn up at the planning stage, it is a living document that must be updated as the research project progresses.

FAIR Principles

European principles for the quality of research data and associated metadata. The acronym FAIR stands for *findable*, *accessible*, *interoperable* and *reusable*. The FAIR Principles guide the drawing up of metadata in particular. Findable means that research data has a unique persistent identifier that functions as a link to the data that can always be found even if the storage location changes. Findability can also be implemented for non-digital research data whose metadata are openly available. Accessible means that research data and the associated metadata are accessible via web browsers. Interoperable means that data are stored using open file formats and common standards. Reusable means that research data has rich metadata and a licence that specifies the terms of reuse.

Information security incident

An event or circumstances that diverge from normal, which may, for example, delay, prevent or harm the conduct of research. The nature of such incidents varies by research field.

Information security incident related to data protection refer to circumstances due to which research data are destroyed, lost, altered or disclosed without authorisation, or due to which an unauthorised party gains access to the data.

Long-term preservation

The preservation of digital data in understandable and usable form for dozens or even hundreds of years. Long-term preservation is designed for valuable research data. The goal of long-term preservation is to ensure the accessibility, authenticity, understandability and completeness of digital objects, also regardless of the eventual obsolescence of or changes to hardware, software and file formats. Long-term preservation ensures the long-term availability of research data.

Metadata

Metadata is data about data. There are several types of metadata, including descriptive, structural, administrative, statistical and legal as well as reference and citation metadata. Metadata ensures the findability and reusability of research data. When research data is described and documented appropriately, other users can trace and understand specific elements of research. Metadata makes it easier to search for and find research data stored in data repositories.

Persistent identifier

A unique and unambiguous, machine-readable name for research output, commonly abbreviated as PID. Identifiers constitute permanent links that always lead to, for example, publications or metadata pages associated with research data. Persistent identifiers enable the long-term findability of digital research data.

Personal data

Any data that can be linked to living natural persons, that is, data associated with or linkable to an identified or identifiable person. Any data that can be used to indirectly identify a person, for example, by linking a specific detail to another detail that would enable identification (pseudonymised personal data) also constitute personal data. Personal data can be stored in, for example, digital files and databases, on paper, in card indexes, in document files and survey forms, or in audio or video recordings.

According to data protection regulation, specific personal data constitute what is known as **special categories of personal data**, or sensitive personal data (*see also ‘sensitive and confidential data’*):

- Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs or trade union membership
- Genetic data
- Biometric data processed for the purpose of uniquely identifying a person

- Data concerning health
- Data concerning a person’s sex life or sexual orientation
- Data concerning criminal convictions or offences

In addition to the data above, the processing of personal data can be sensitive if the processing may pose risks to the data subjects (*see ‘risk management’*).

Research data

Data that have been collected, observed, measured or created to verify research results, or that are otherwise considered necessary in the research community for the confirmation of results. The context turns data into research data. Any data can become research data when it is analysed for research purposes.

Research data can include measurement results, test results, survey results, audio and video recordings as well as samples and specimens. While research data is often in digital form, it can also include physical or analogue data. Research data can be raw data, processed data, data in the possession of a third party, shared data or published data. The degree of openness of research data varies from confidential and sensitive data to open data.

Research data infrastructure

Processes, technical solutions and services through which research data management is carried out in practice. Research data infrastructures involve organisation, an operating culture and long-term social networks that enable the realisation of technical and administrative solutions and services.

Research data management

Commonly referred to using the abbreviation RDM. A process encompassing the lifespan of the research project that includes the collection or acquisition, organisation, curation, storage, (long-term) preservation, protection, quality assurance, licensing and distribution of research data as well as the use of persistent identifiers and other metadata in compliance with the rules and procedures of the relevant discipline (European Commission/Horizon Europe).

Research infrastructure

Research infrastructures are instruments, equipment, information networks, databases, materials and services that serve to facilitate research, promote research collaboration and reinforce research and innovation capacity and know-how (Academy of Finland).

Risk management

Proactive anticipation of events that have negative consequences. Risk management constitutes coordinated activities used to guide, manage and monitor the actions of the University, its units or individual researchers in relation to risks. The purpose of risk management is to help the University, its units and researchers to attain their goals and make decisions.

When processing personal data (*see the definition of 'personal data'*), the risks associated with the processing must be assessed while ensuring the implementation of the relevant data protection principles. In the case of personal data, the risk assessment must be carried out from the perspective of data subjects.

Sensitive and confidential data

Research data whose storage, use and sharing are restricted on ethical, legal, contractual or commercial grounds. Such data must be processed and protected with particular care.

Sensitive data are associated with, for example, specific personal data (*see special categories of personal data in the definition of 'personal data'*), endangered species, biosecurity or national defence. Confidential information on patents and trade secrets are also to be protected, and their exposure can result in claims for damages.