

The Research Council of Norway's Policy for Open Access to Research Data

December 2017

Content

| | |
|---|---|
| Foreword | 1 |
| 1. Introduction | 2 |
| Objectives | 2 |
| The role of the Research Council | 2 |
| Definitions and delimitations | 3 |
| Use of data in the research process | 3 |
| 2. Principles and guidelines in the Research Council's policy | 4 |
| Main principle | 4 |
| Exceptions to open access requirements | 5 |
| The Research Council's guidelines | 6 |
| Clarification of terms | 7 |

Foreword

Transparency and knowledge sharing are critical for research. Rapid technological development is changing how research is conducted and results are shared. Advancement in research is increasingly being driven by access to new, large amounts of data. The results of publicly funded research comprise a public good that is valuable to researchers as well as to society at large. Better access to research data will enhance the quality of research as the results can be validated and verified more efficiently and data sets can be used and combined in new ways. Creative and extended use of research data may pave the way for more wide-ranging interdisciplinary research, as well as innovation of societal and industrial value. In addition, data sharing may help to prevent duplication of efforts and facilitate the use of data sets in other contexts than originally planned. Open access to results and data will also be valuable for education and give research a higher profile within society at large.

This policy sets out how the Research Council of Norway (from now on referred to as the Research Council) will use its instruments to promote open access to research data. The principles and guidelines presented in this document are intended to serve as a guide for researchers when planning their research projects.

This policy was published in 2014 and will be revised in line with experiences gained from its implementation, as well as international developments, changes to the legal framework, and new technological potential. The basis for this update (2017) is related to the growing number of research funders, including the EU, that are requiring R&D projects to submit a data management plan describing how the research data in a project will be managed from the time it is collected until after the project ends. In line with this, the Research Council has introduced new guidelines for data management in 2017. The principles and guidelines presented in this policy conform to the *Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata* [“National Strategy on access to and sharing of research data”, currently Norwegian only] issued by the Government in December 2017. The main principles of the policy from 2014 remain for the most part unchanged, but the following additions have now been included:

- The Research Council has introduced guidelines requiring that R&D-performing institutions or companies should assess whether projects receiving funding from the Research Council must develop a data management plan. Furthermore, the policy establishes that data management plans ought to be publicly accessible, and be made openly accessible by the R&D institution. This has the potential of promoting greater openness and enabling scientific groups to follow and learn from peer practice.
- The FAIR Guiding Principles for scientific data management and stewardship are included as a main principle in the Research Council’s policy. FAIR is an acronym for the words *findable, accessible, interoperable* and *reusable*.
- The international FAIR Principles (see page 4) for use of research data have been incorporated into the main principle of the policy.
- Several of the guidelines listed on page 5 have now been reworded from recommendations to requirements in order to ensure compliance with the FAIR principles for reuse of research data.

1. Introduction

Objectives

The Research Council has established this policy to lay the foundation for improved quality assurance, storage, accessibility to and sharing of data. This will lead to:

- Enhanced research quality through the ability to expand on previous studies and combine data in new ways;
- Transparency in the research process and better opportunities for verifying scientific results;
- Increased cooperation and less duplication of research efforts;
- More innovation in the private and public sectors;
- Greater efficiency and better utilisation of public funding.

The role of the Research Council

Storing and making research data accessible for further use must take place in a cooperative effort between researchers, research institutions, scientific journals, infrastructure providers and

government authorities. Achieving this will require development of new competency and training as well as investment in new infrastructures and tools. The Research Council seeks to be a driving force for more widespread storage and sharing of research data. This means that the Research Council will facilitate cooperation between the stakeholders, finance relevant activities, and provide guidance to the research communities by:

- Investing in infrastructure that supports quality assurance, long-term storage, sharing and use of research data;
- Promoting an effective division of tasks between public infrastructure and service providers through the facilitation of meeting places and the provision of advisory services;
- Continuing to require through the Research Council contracts that research data are stored in a safe and secure manner for a minimum of 10 years;
- Encouraging the institutions to develop their own policies and guidelines for secure storage and management of, and improved accessibility to, research data;
- Establishing a framework that enables R&D-performing institutions or companies to assess the need to develop a data management plan (DMP) for projects that receive funding.

Definitions and delimitations

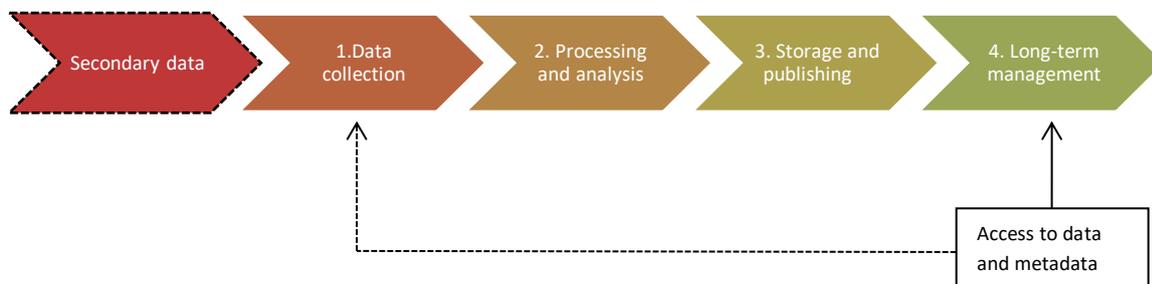
The principles and guidelines in this policy apply only to research data that are generated through research activity funded wholly or partly by the Research Council. The policy focuses on research data that are accessible in a digital format.

The term “*research data*” is defined in this policy to mean the registration/recording/reporting of numerical scores, textual records, images and sounds that are generated by or arise during research projects. These may, for example, be data that are generated through new analysis by combining existing secondary data, or entirely new data that are generated through new data collection. Research data are always a direct result of research activity, regardless of whether the data are based on secondary data or whether they are collected from scratch.

The term “*secondary data*” is used in this policy to refer to data that already exist, regardless of the research to be conducted. These may comprise information collected for a different purpose (e.g. public administrative data, clinical data or weather data) or they may be physical or digital collections of objects and texts (e.g. libraries or data reused from previous projects such text corpuses or other scientific collections). Information on the Internet may also be defined as secondary data in this context, and such information is highly heterogeneous. Data which are used as secondary data in research, but which have been collected, generated or processed by other researchers or research institutions than those conducting the research, will normally not be encompassed by the guidelines in this policy.

Use of data in the research process

To enable the reuse of research data, the data must be processed after they are generated. The value chain for digital research data is divided into four main steps:



1. The first step is the collection or production of data, where the data are stored in a workspace. It is best practice to employ accepted documentation standards already at this stage, as it will simplify subsequent efforts relating to long-term storage and publication.
2. In the second step, the data are quality assured. This is a process for determining whether the raw data are to be preserved. Analysis of the quality-assured data is also carried out during this phase.
3. The third step is preparing the data for long-term storage (archiving) and publication. This involves discipline-specific documentation and coding of data (metadata) describing i.e. who is responsible for the data, what the data contain, what has been done with the data, who can use them for which purpose. These metadata will make it possible for other researchers to find and reuse the data. Relevant information about the data will be indexed in search engines/catalogues when the data are transferred to a data repository. As the data are stored in a data repository, the data set is assigned an identifier which follows the data set throughout its lifetime. The data are now accessible to other researchers.
4. When the data are stored in a repository they are preserved for the future. This usually entails a storage period of minimum ten years or more. Technologies change, which means that data deposited in long-term storage must be adapted to new technologies and formats to ensure that they remain findable, available, interoperable and reusable in the future.

2. Principles and guidelines in the Research Council's policy

Main principle

Open access

The Research Council's policy follows an open-by-default principle regarding access to research data. The Research Council will therefore help to ensure that research data in general are made openly accessible, but that exceptions are made for data that cannot or should not be openly accessible. The

Horizon 2020 programme's definition¹ of open access states that access to scientific information must be, as far as possible, free of charge. The Research Council has decided to base its policy on the premise that the user should cover the actual costs incurred from data retrieval. This is closer to the OECD's definition² of open access, which states that access must be provided at the lowest possible cost, preferably at no more than the marginal cost of dissemination.

The FAIR Principles for sound data management

The international FAIR Principles have been formulated as a set of guidelines for the reuse of research data. The acronym FAIR stands for *findable, accessible, interoperable and reusable*.³ Research data must be of quality that makes them accessible, findable and reusable. The concept *interoperable* entails that both data and metadata must be machine-readable and that a consistent terminology is used.

Exceptions to open access requirements

There are a variety of challenges associated with making some data sets openly accessible. The reasons for restricting access include:

Security concerns

When open access to the data may threaten personal or national security, the data sets **must not** be made openly accessible.

Sensitive personal data

When open access to the data conflicts with the applicable statutory framework regarding the protection of personal privacy, the data sets **must not** be made openly accessible.

Other legal factors

When open access to the data conflicts with other legal provisions, the data sets **must not** be made openly accessible.

Commercial factors

Data that have commercial value and are generated in projects in which a company is the contractual partner with the Research Council **may** be exempted from the general principle of open access. In these cases, it is recommended that the data are made available after a certain period of time, preferably after three or five years.

¹ European Commission: "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020", Version 3.2, 21 March 2017 [Downloaded 20.11.2017].

² OECD: "Principles and Guidelines for Access to Research Data from Public Funding", April 2007 [Downloaded 20.11.2017] and OECD: "Making Open Science a Reality" 2015. OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris.

³ Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016). <https://doi.org/10.1038/sdata.2016.18>. See also: <https://www.force11.org/group/fairgroup/fairprinciples>

Other factors

When open access to data will have major financial or practical implications for those who have generated/collected the data, the data sets *may* be exempted from the general principle on open access if a satisfactory argument is made for this.

The Research Council's guidelines

The Research Council has formulated a set of guidelines for the storage, dissemination and sharing of research data. The guidelines will be followed up through the Research Council's research funding instruments and through the National Financing Initiative for Research Infrastructure.

- 1.0 Research data must be stored/archived in a safe and secure manner.
 - 1.1 The data must be stored in secure archives, either in a trusted repository at the relevant institution or in national archives.
 - 1.2 The research-performing institution is responsible for selecting relevant archiving solutions for the various research data in accordance with the FAIR Principles.
 - 1.3 The projects should explicitly address how to manage research data that are considered to have long-term value.
 - 1.4 The projects should explicitly address how research data that are not considered to have long-term value should be managed, or if relevant destroyed, after a certain period of time.
- 2.0 Research data must be made accessible for reuse.
 - 2.1 Research data must be made accessible to all relevant users, on equal terms, as long as there are no legal, ethical or security-related reasons to preclude this.
- 3.0 Research data should be made accessible at an early stage.
 - 3.1 The data used as the basis for scientific articles should be made accessible as soon as possible, and never later than at the time of publication.
 - 3.2 Other data that may be of interest for other research should be made accessible within a reasonable amount of time, and never later than three years after the project has ended.
- 4.0 Research data must be accompanied by standardised metadata.
 - 4.1 The metadata must enable others to search for and use the data.
 - 4.2 The metadata should follow international standards.
 - 4.3 The metadata should describe the quality of the data.
- 5.0 Research data must be provided with a license for access, reuse and redistribution.
 - 5.1 The license should be internationally recognised.
 - 5.2 The license should set as few restrictions as possible on the access, reuse and redistribution of the data.
- 6.0 Research data should be made accessible at the lowest possible cost.
 - 6.1 Metadata should be made accessible at no charge and be published in a way so that they may be harvested and used in research data searches.
 - 6.2 Research data should preferably be made accessible at no charge.
 - 6.3 The price of access to research data should never be higher than the actual costs of making them available.
- 7.0 The management of research data must be described in a data management plan.
 - 7.1 The research-performing institution must approve that data management plans follow the institution's requirements and expectations. The plan should comply with the FAIR Principles.

7.2 As far as possible, data management plans should be public and made openly accessible by the research-performing institution to enable scientific groups to follow peer practice.

Clarification of terms

Some of the key terms related to open access to research data can be ambiguous. Consequently, it is important to specify what the Research Council means by these terms.

| | |
|---------------------------------|--|
| Research data | <i>is understood in this policy to mean the registration/recording/reporting of numerical scores, textual records, images and sounds that are generated by or arise during research projects.</i> |
| Secondary data | <i>are data that already exist, regardless of the research to be conducted.</i> |
| Publicly funded | <i>refers to all projects and activities that are wholly or partly funded with government allocations.</i> |
| Open access | <i>is understood as the principle that research data should be accessible to relevant users, on equal terms, and at the lowest possible cost. Access should be easy, user-friendly and, if possible, Internet-based.</i> |
| Data management plan | <i>is a document describing how research data from a project are to be managed, from project start to finish.</i> |
| Metadata | <i>are data used to define or describe other data.</i> |
| Digital Object Identifier (DOI) | <i>is a persistent, unique identifier for research data and publications.</i> |