NWO Institutes Data Management Policy Framework

Introduction

NWO's strategy for open science strives for optimum accessibility of the results from the research it funds. The NWO Institutes Data Management Policy Framework is an elaboration of the wider NWO strategy in the area of Open Science, which also includes Open Access of scientific publications and Scientific Integrity. An Open Access policy has already been developed and implemented for publications, for example via the NWO Regulation on Granting. In the NWO grant programmes, applicants are requested to write a data section in their proposal and if the proposal is awarded funding then a data management plan must be written as well.

This policy framework has been written for the institutes to assist them in responsibly handling and managing research data with a view to both the replication of the research and the reuse of the research data for new research. The institutes will elaborate this framework further.

In 2015 an input memorandum was written, in collaboration with the data managers from the institutes and external experts, for the elaboration of this Policy Framework Data Management NWO Institutes. In June 2016 the institute directors approved this framework. Subsequently the Governing Board of NWO approved this policy framework in its meeting of 24 August 2016.

The aim of this policy framework

- The policy framework provides guidelines for handling and managing research data, while still allowing the institute to take domain-specific requirements into account.
- The policy framework intends to facilitate the reuse of research data and to enable the replication of research.

Framework for data management policy at institute level

Responsible handling and careful management of data are, of course, part and parcel of scientific research. These responsibilities are sometimes also described as data stewardship.

Data management policy at the institute level aims to ensure that research data are findable, accessible for the long-term, and can be reused, either independently or in combination with other research data. The data process should be transparent and of such a quality that the research data can be replicated.

As such, NWO subscribes to the internationally accepted FAIR principles as the standard for managing research data. This means that research data must be easy to find, accessible, interoperable and reusable. Furthermore, the data should be archived in a safe environment.

Starting points:

- Optimal use of research data is central to the definition of the data management policy.

- The policy framework is not an assessment framework but should serve as a guideline for the development of a data management policy at the institute level.

- The NWO institutes will of course be given the freedom to set up their data management policy in accordance with domain-specific requirements and the international practice within their scientific field.

- The FAIR principles and the safe preservation of data form the starting point for the management of the research data.

Guidelines for the data management policy of the NWO institutes are given below.

Long-term and safe storage
- The institute is responsible for the storage of research data and the associated metadata during and after the research process.
- Once the research process has been completed, the research data will be preserved in an archive; in view of the current legislation the minimum storage period is 10 years. The storage period starts at the moment that the research process has been completed.
- The institute describes in which repository the data will be archived. That can be the institute's archive or a 'trustworthy digital repository'.

Findable
- The digital research data must be, and remain to be, easy to find in the archive. A persistent identifier will be created and/or requested from a archiving institution, to increase the findability and to ensure that the data is adequately referred to.
- The replication package[1] will be offered to 'harvesters' (via the archive) in such a way that the data are findable via standard (domain-specific) portals.

Access
- If conditions are attached to the access, then these must be made known by the institute. In such cases the institute describes the conditions for access and reuse, and states which selection criteria will be used for granting access to the data. The default position is, of course, free access to data.

Interoperable
- The metadata will be aligned with the accepted national and international standards within the scientific discipline.
  If a broadly accepted standard is lacking within the discipline, then connections can be sought with the best practices of other institutes and with the relevant developments in the field during the development and implementation stage of the data management policy.

Reusable
- The institute ensures that a replication package is provided with information about the data.
- The metadata will be drawn up in accordance with the domain standard and if this is lacking then a widely accepted generic standard for metadata will be sought and used.

Replicable
- The archived data, metadata, and replication package, taken together, must be sufficient to replicate the published research.

[1] The term replication package is defined on page 4 of this document under the heading 'Meaning of the terms as used in this policy framework'.

- The institute assumes responsibility for a description of the required (domain-specific) content of the replication package, satisfying the above condition.
- In response to a reasoned request from a researcher or another interested party, the institute will also make the associated logbooks, journals, et cetera available in addition to the replication package.

Data stewardship: responsibilities

Data stewardship is the careful and responsible handling and management of research data both during and after the research.

The roles and responsibilities with respect to data stewardship for the parties involved are described below. Any deviations from the guidelines below in the fulfilment of roles and responsibilities due to domain-specific agreements will be stated in the institute policy.

Researchers

- The researcher ensures, in consultation with the data manager of the institute, that the research data and the associated metadata are safely stored both during and after the research process: in a storage facility provided by the institute, and/or as part of a national or international data archive.
- In consultation with the data manager of the institute, the researcher will provide a replication package for the research data.

Institute director

- The institute director is responsible for drawing up, implementing, monitoring and evaluating the data management policy.
- The institute director is ultimately responsible for the scientific research at the institute and the institute director retains the ownership of and control over the research data and the metadata, on the institute's behalf.
- The institute director can delegate responsibilities but always bears the final responsibility.

Institute board

- The institute boards ensure that the data management policy is drawn up and implemented at the institute level.

NWO Executive Board

- The NWO Executive Board monitors the timely and adequate implementation of the data management policy at the research institutes.

National context of the policy framework

Dutch legislation applies to this policy framework. The legislation referred to below is not an exhaustive summary but instead serves to illustrate the most important legislative frameworks and how these relate to data management.

- Codes of conduct[2]
  - Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek
  - Code for Transparency in Animal Testing
  - The Netherlands Code of Conduct for Academic Practice (VSNU)

[2] In the context of data management these codes are not only important for how the data is collected and stored, but also concern the possibility of replicating the research and the minimum period for the preservation of research data.

- Dutch Copyright Act and the Dutch Patents Act
  The rights to publications of scientific research results and the underlying data are described in this.
- Dutch Databases Act
  The rights to the dataset and positions of the funder and producer of the databank are described in this.
- Dutch Personal Data Protection Act
  This is relevant in light of the restrictions that can apply to sharing and reusing research data.
- Collective Labour Agreement Research Institutions
  The intellectual property rights are described in this (Art. 1.9) – which are also applicable to the databanks produced.

Meaning of the terms as used in this policy framework

Data can be: facts, observations, interviews, recordings, measurements, experiments, simulations and software; numerical, descriptive and visual; raw, cleaned up and processed; whether or not to support an actual or intended publication; and stored and exchanged in various formats on various storage media.[3]

Data management is understood to include the entire pathway from the creation or collection of data to the storage, maintenance, archiving, disclosure and long-term storage (preservation) of data. No distinction is drawn between the aims of data storage such as checking, verification, replication, reuse or linking of the data.[4]

Metadata - metadata is information about data. Generic and domain-specific standards exist for metadata; the domain-specific standards often provide richer descriptions but are not always supported in broad portals.

Persistent identifier - a worldwide unique code that identifies digital objects such as a dataset or a publication. Unlike ordinary bookmarks a persistent identifier also continues to refer to the object if this is relocated. A persistent identifier can therefore be used to consistently cite data and publications.

Provenance or in other words origin of the research data. Dependent on the discipline, this can be found, for example, in logbooks and lab journals, research protocols, data management plans, instrument configurations, database queries, reused information from data repositories, and contracts with data suppliers.

Replication package - the full set of data, metadata including the persistent identifier and provenance information, documentation, possibly a description of the required software, hardware and tools, as well as a reference to where the log books, lab journals, research protocols et cetera can be consulted. This package of information is archived together with the data. It is the set of details and information needed to be able to reuse and replicate the data.

[3] This is in line with the definition used in the Berlin Declaration (2003); https://openaccess.mpg.de/Berlin-Declaration.

[4] This definition is based on the report “Inventory Data Management NWO institutes” (March 2014).