

Research Data Management

Institute for Computing and Information Science

1. Preliminaries

This document is an addition to the policy of the Radboud University on storage and management of research data (Executive Board decision dated 25-11-2013, see www.ru.nl/rdm/vm/policy-documents/).

Research data entails any information gathered or generated in the process of executing a research project; this includes research designs, questionnaires, lab notes, etcetera. In the specific case of computing science, a major information source produced or studied in the research is software, the source code of which is then the research data of concern.

The university wide policy can be summarised in three main points:

- Research data need to be archived for a minimum term of ten years, ultimately from the first moment of publication of research based on this data, together with sufficient information to enable reuse.
- A data management plan should be written, preceding data collection.
- The primary responsibility to take care of research data management is the individual researcher's (with ultimate responsibility for the director of the institute).

2. Institute policy objectives

The primary goal of the institute's Research Data Management (RDM) policy is to *implement a research practice that leads to reproducible research results*. Hereto, iCIS follows the FAIR Guiding Principles for scientific data management and stewardship (Nature, March 2016):

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

iCIS strives to realize a situation where the default rule is that published iCIS research can be independently reproduced. Researchers therefore have to publish, along with their research publications, the data and the software required to carry out such a reproducibility study.¹

A second objective of the institute's RDM policy addresses the prevention of the loss of research data, for example due to a crashing/corrupted external hard disk or the physical loss of portable storage media (e.g., a USB drive or a laptop). In this case, the policy addresses activities during the research itself, and not just upon completion.

The third and final objective of this policy is to help prevent unauthorized access to research data, in particular when research involves personally identifiable information.

3. iCIS Data Management Protocol

The iCIS Data Management Protocol details the practical implementation of RDM in the research workflow, including requirements regarding a Data Management Plan per project as well as the way data and code is stored and documented.

The Data Management Protocol is specified in a separate document, targeting the institute's researchers (junior and senior alike). In brief, the protocol prescribes 1) the existence of a Data Management Plan for every project, 2) the default use of version control for all research related outcomes (publication, data and code), and 3) the archival of a snapshot of the corresponding repository in RIS upon publication of research results.²

Link to current draft protocol:

<https://www.ru.nl/icis/research-data-management/policy-protocol/>

4. Roles and responsibilities

Researchers, data steward and institute director implement together the iCIS RDM policy to realize the shared goal of reproducible research results:

- iCIS research staff familiarise themselves with the RDM protocol, and carry out their research in compliance with the protocol;
- The data steward monitors compliance with the data management protocol by sampling, on a regular basis (once every quarter), published research papers and theses, and checking the archival status of their underlying data and code;
- In case of non-compliance, the data steward will inform the institute director. Together they will decide which action is necessary to address this non-compliance.

5. Further reading

The Research Information Services (RIS) desk supports a highly recommended website www.ru.nl/research-information-services/, structured around the phases of the data lifecycle from planning research to archiving data.

¹ Note that it is unlikely that *all* data would be published as *open data*; nor will all software be publically available as open source code to anyone on the internet. An independent third person with proper access rights should however be able to reproduce the findings reported in the publication.

² A research project's Data Management Plan could motivate an exception for (potentially large) data sets, for example when the source data resides in a reputed disciplinary repository (e.g., a [TREC collection](#), a data set from the [UCI Machine Learning Repository](#), etc.).