

1. A broad introduction to Research Data Management

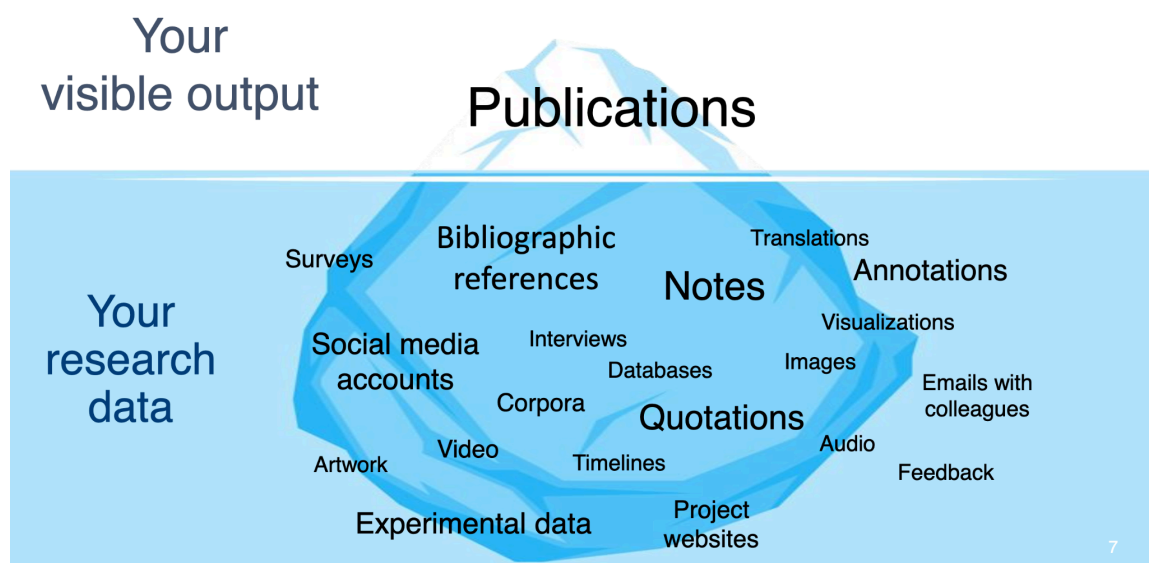
For a quick introduction, watch the following video: Ghent University Data Stewards (2020).

Knowledge clip: [What is Research Data Management \(RDM\)?](#)

Why do we talk about research data?

Publications are just the visible output of research. Most of the work of scholarship happens within the vast amount of materials that underlie each publication: the research data.

This introduction explores ... XXX Add brief overview of the text here (exec summary)



Source: Noppe, N., Vanvelk, J., & Callens, N. (2021). The hands-on guide to research data management for KU Leuven researchers, students, and research support staff in the humanities and social sciences. [Zenodo](#)

Responsible conduct in research data management: core values

Research integrity

Research integrity means ensuring that all research is conducted in a way that is trustworthy and reliable. It applies to every field and discipline of research, and to all sectors in which research activities are performed.

Research integrity covers all phases of the research process, from the initial idea and planning of a research project to carrying it out and sharing the results. Research integrity also involves ensuring that research environments and systems support and improve good research practices, often referred to as "research culture."

Standards of conduct for scientific integrity are laid out in codes of conduct, such as:

The Netherlands Code of Conduct for Research Integrity, 2018 "If scientific and scholarly research is to perform this role properly, research integrity is essential. Researchers who are not guided by the **principles of honesty, scrupulousness, transparency, independence and responsibility** risk harming both the quality and the trustworthiness of research."

The European Code of Conduct for Research Integrity, Revised edition 2023 "Good research practices are based on **fundamental principles** of research integrity. They **guide individuals, institutions, and organisations** in their work as well as in their engagement with the practical, ethical, and intellectual challenges inherent in research."

These fundamental principles include:

Reliability in ensuring the quality of research which is reflected in the design, methodology, analysis and use of resources.

Honesty in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair, full, and unbiased way.

Respect for colleagues, research subjects, society, ecosystems, cultural heritage and the environment.

Accountability for the research from idea to publication, for its management and organisation, for training, supervision and mentoring, and finally for its wider societal impacts.

★ In order to prevent research data from being invented, fabricated or falsified (intentionally manipulated or misinterpreted) and to stimulate reproducibility, a **transparent researcher clarifies on what data he/she bases conclusions and how these data were collected.**

Reproducibility

Research is considered to be reproducible when researchers provide all the necessary data, metadata, software and computer scripts enabling the research to be reconstructed or making it possible to re-analyse the research data.

[! Note] "Reproducibility means that research data and code are made available so that others are able to reach the same results as are claimed in scientific outputs" | [The Open Science Training Book, 2018](#)

Setting up reproducible research requires extensive preparation in the planning phase. Research data which are not accurately managed risk not being reusable, which lowers the possibilities of reproducing the research itself.

Data management vocabulary

This section should clarify the meaning of some of most commonly used terms in research data management activities

Research Data

The meaning of the research data heavily depends on the scientific discipline. Below are provided a few definitions which outline the main features and ways of considering research data.

OECD, 2007 "Research data are the **factual records** (numerical scores, textual records, images and sounds) **used as primary sources for scientific research**, and that are commonly accepted in the scientific community as **necessary to validate research findings**. A research data set constitutes a systematic, partial representation of the subject being investigated."

Utrecht University, 2016 "Research data constitute **primary research data** (the raw, rough measurements or observation) and **secondary research data** (the results after the data have been processed by a researcher (recoded, combined, categorised, visualised, etc.))."

Van Berchum & Grootveld, 2017 "Research data may be **facts, observations, interviews, recordings, measurements, experiments, simulations and software; numerical, descriptive and visual; raw, cleaned up and processed**; they **may or may not support an actual or intended publication**; and may be **stored and exchanged in various formats on various storage media**."

Research data can be: - information collected or produced in the course of scientific activity. - considered as evidence needed to validate research conclusions and results, and can be classified in various ways. - different between raw/primary data, i.e. the direct result of measurement or collection, and secondary data, when reusing data produced by others. - numeric, textual, or multimedia, for instance, statistics, digital image collections, sound recordings, interview transcripts, etc. are considered data.

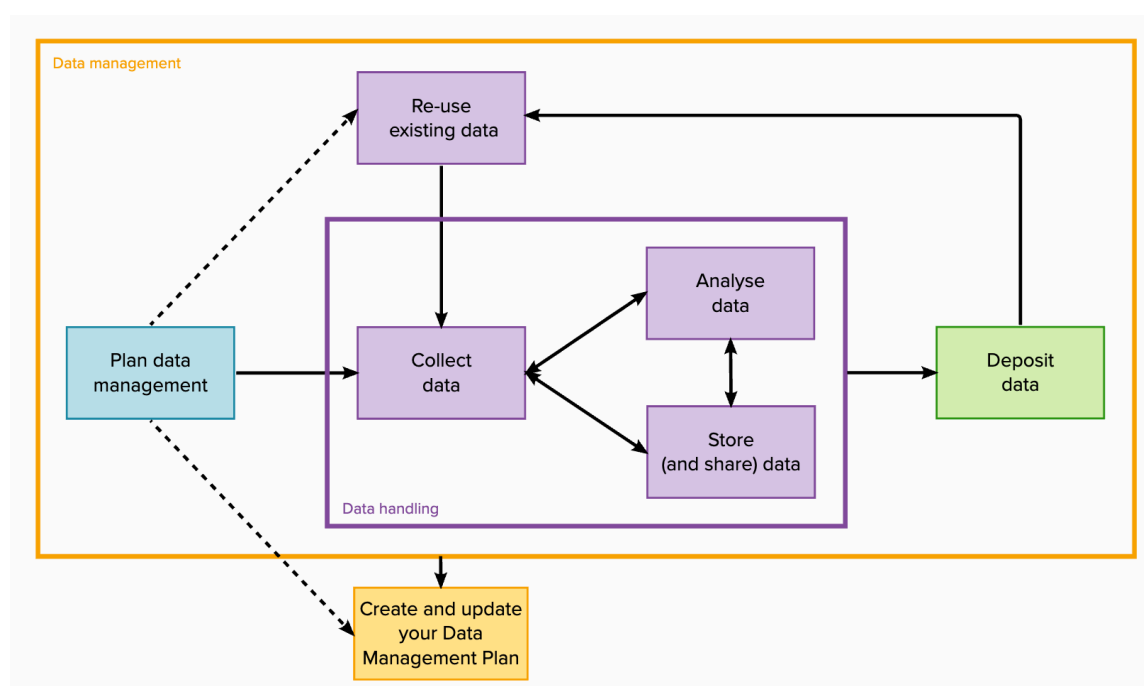
So when we talk about 'research data' we are using a generic term that can refer to a variety of information and contents, expressed in different formats, and that can be creative or

factual in nature.

Research Data lifecycle

One of the ways of looking at research data is according to the phase of the research process, thinking about what you intend to do with the data in a structured manner.

Following the research lifecycle model, the data lifecycle is used to represent how the various phases of the life of research and research data are combined, and how the choices made in one stage can affect the quality of data in another.



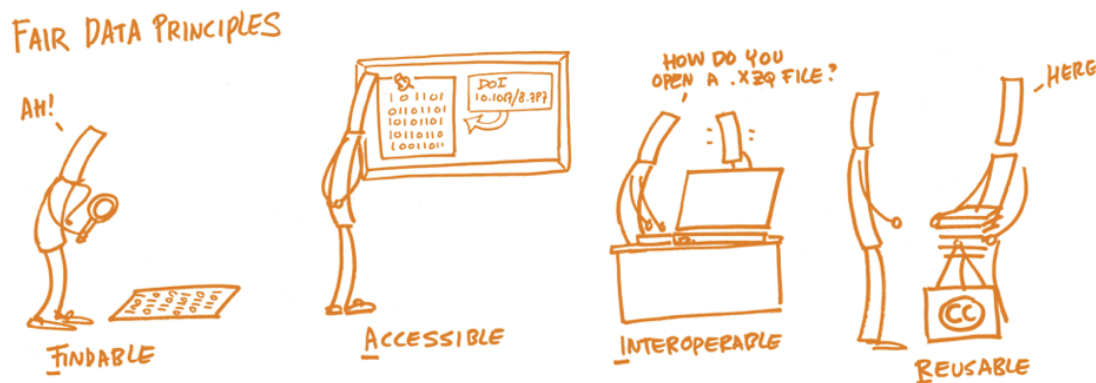
Source: Gualandi, B., Caldoni, G., & Marino, M. (2022). Research Data Management: Data Lifecycle. [Zenodo](#).

Plan data management: description of the data that will be compiled, making clear how the data will be managed and made accessible throughout its lifetime. **Re-use** and/or **Collect data:** corresponds to the data collection phase, and may include primary/raw and processed data. **Analyse data:** the phase in which you handle and work on your data. **Store (and share) data:** data is described appropriately, following standards when they exist. **Deposit data:** ensure that data is accessible in the future, in particular that it is stored in such a way that others can use it, using a data repository and following FAIR principles.

FAIR principles

The [FAIR Guiding Principles for scientific data management and stewardship](#) are a set of principles for the management of research data which were defined by a group of stakeholders from academia, industry, funding agencies, and scholarly publishers between 2014 and 2016. The principles describe how data should be managed in order to be **FAIR: Findable, Accessible,**

Interoperable and Reusable. FAIR Data is therefore data which has these four characteristics and has been identified by the **European Commission** as one of the eight pillars to promote the uptake of Open Science.



Source: Martínez-Lavanchy, P.M., Hüser, F.J., Buss, M.C.H., Andersen, J.J., Begtrup, J.W. (2019). 'FAIR Principles'. In: Holmstrand, K.F., den Boer, S.P.A., Vlachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.) (2019). *Research Data Management* (eLearning course). doi: 10.11581/dtu:00000047

There are in total 15 principles of a technical nature, from which derive a set of best practices which come into play throughout the research data lifecycle: from planning to collection, from analysis to sharing, and from preservation to reuse.

For each letter of FAIR, sub-principles provide technical details.

Findable

Others can easily discover your data

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

Accessible

It is clear who, when and how can access your data (does not mean open!)

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

Interoperable

Your data can be integrated with other data and/or they can be easily used and read by machines

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

Reusable

Your data can be reused by others in new research

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes R2. (Meta)data are released with a clear and accessible data usage license R3. (Meta)data are associated with detailed provenance R4. (Meta)data meet domain-relevant community standards

[!NOTE] The FAIR principles can be understood as the requirements that research data and results should have in order to adhere to the model of Open Science, in order to make it possible to trace these data within scientific production and to facilitate reuse, when possible, in the creation of new knowledge. You will often hear about FAIR Data, but it is also possible to apply the FAIR principles to other research outputs such as presentations, and publications.

Open Data

tbd

Research Data Management

tbd: questions to ask yourself to look at your research data

Data stewards