

Glossary Flemish Standard Data Management Plan

Accession number: A systematic (computer-readable) number or code that uniquely identifies an entry in a particular database¹. Accession numbers are commonly used in the Biomedical and Life sciences.

ARK: Archival Resource Key

Dataset: Data and objects generated or collected by researchers in the course of their investigations, regardless of their form or method, that form the object on which researchers test a hypothesis. This includes the full range of data: raw, unprocessed datasets, proprietary generated and processed data and secondary data obtained from third parties².

Data format: The way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf)³.

Data usage license: A data usage license indicates whether the data can be reused or not and under what conditions. If no licence is granted, the data are in a grey zone and cannot be legally reused. Do note that you may only release data under a licence chosen by yourself if it does not already fall under another licence that might prohibit that.

Data types: Data are often grouped by type (observational, experimental etc.), format and/or collection/generation method.

When focussing upon the mode of data collection, one distinguishes between data created via⁴:

- Observations

These data are captured in real-time, either by human observation and surveys, or instruments or sensors. For this reason, they are usually irreplaceable and most important to store safely. Examples: sensor readings, survey results, audio and/or video recordings of interviews.

- Experiments

These data are typically generated in the laboratory under controlled conditions. They often are reproducible, but this procedure can be expensive or time-consuming. Examples: gene sequences, chromatograms, magnetic field readings.

- Derivation or compilation

These data are generated by combining multiple existing datasets. Examples: text and data mining, compiled database.

- Computations, models or simulations

¹ <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095346934>

² Flemish Open Science Board Metadatamodel for research datasets

³ Science Europe RDM Practical guide, p.18

⁴ <https://bibliotheek.uhasselt.be/en/collect>

These data are machine-generated from test models. The output files are likely to be reproducible as long as the model and inputs are preserved. Therefore, the large-volume output files can often be discarded when wrapping up your research project and selecting data for long-term preservation. Examples: climate models, economic models.

Documentation:

Documentation for data is needed to understand data and enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on. This information can be captured and recorded in different ways, for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks. Documentation can thus be very rudimentary (a README-file provided with the data that generally describes it) or highly specific (structured metadata as generated by specific measurement instruments).

DOI: Digital Object Identifier

A name (not a location) for an entity on digital networks. It provides a system for persistent and actionable identification and interoperable exchange of managed information on digital networks. A DOI is a type of Persistent Identifier (PID) issued by the International DOI Foundation. This permanent identifier is associated with a digital object that permits it to be referenced reliably even if its location and metadata undergo change over time. SYNONYM. DOI⁵

Dual use: "Dual use items" are items, including software and technology, which can be used for both civil and military purposes. The term "misuse" is defined as "research that could be misused for unethical purposes"⁶.

Ethical aspects in research data management may include:

- processing of 'special categories' of personal data (formerly known as 'sensitive data')
- processing of personal data concerning children, vulnerable people or people who have not given their consent to participate in the research
- complex processing operations and/or the processing of personal data on a large scale and/or systematic monitoring of a publicly accessible area on a large scale
- data processing techniques that are invasive and deemed to pose a risk to the rights and freedoms of research participants, or techniques that are vulnerable to misuse
- collecting data outside the EU or transferring personal data collected in the EU to entities in non-EU countries⁷

Intellectual Property Rights: The most common intellectual property rights are those protecting a (technical) invention, a trademark, a new plant variety, (industrial) designs and

⁵ <https://codata.org/rdm-glossary/digital-object-identifier/>

⁶ <https://bibliotheek.uhasselt.be/en/ethical-and-legal>

⁷ https://ec.europa.eu/info/sites/default/files/5_h2020_ethics_and_data_protection_0.pdf

literary/artistic works. If you are collaborating with external partners, it is important to clarify who owns the IPR of the creations resulting from the research⁸.

Open access: The practice of providing online access to research outputs, free of charge to the end-user, and without any legal or technical obstacles, such as the requirement to have a user account or to solve a captcha⁹.

Metadata: Metadata are data about data. Research data need metadata to become findable, accessible, interoperable and reusable - by humans and machines. Metadata play an important role in making your data FAIR. Metadata have to be added continuously to your research data, not just at the beginning or at the end of a project. Metadata can be added manually or automatically, and preferably according to a disciplinary standard. From a FAIR perspective, metadata are more important than your data, because metadata would always be openly available and they link research data and publications in the Internet of FAIR Data and Services. The distinction between data and metadata is not ontological, but it is grounded in use. What is “data” and what is “metadata” is thereby a matter of perspective: Some researchers’ metadata can be other researchers’ data¹⁰. While data documentation is meant to be read and understood by humans, metadata (which are sometimes a part of the documentation) are primarily meant to be processed by machines¹¹. For more information please visit: <https://www.howtofair.dk/how-to-fair/metadata/>

ORCID: Open Researcher and Contributor ID¹²

Physical, non-digital, or analogue or data or research materials:

Physical data are equally considered research data. Obviously, these data require a completely different approach regarding, for example, storage and preservation. Examples: paper-based questionnaires and notes, archaeological findings, art works (e.g. paintings, sculptures, photographs), protein and blood samples, nucleic acids, building plans, recordings on tapes or discs¹³.

Personal data:

Personal data is any information that relates to an **identified or identifiable living individual**. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data.

Personal data that has been de-identified, encrypted or **pseudonymised** but can be used to re-identify a person remains personal data and falls within the scope of the GDPR. Personal data that has been rendered **anonymous** in such a way that the individual is not or no

⁸ <https://bibliotheek.uhasselt.be/en/ethical-and-legal>

⁹ <https://erc.europa.eu/managing-your-project/open-science>
https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/open-access_en

¹⁰ <https://www.howtofair.dk/how-to-fair/metadata/>

¹¹ <https://www.howtofair.dk/how-to-fair/metadata/>

¹² <https://orcid.org/>

¹³ <https://bibliotheek.uhasselt.be/en/collect>

longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible¹⁴.

Examples:

- a name and surname;
- a home address;
- an email address such as name.surname@company.com;
- an identification card number;
- location data (for example the location data function on a mobile phone)*;
- an Internet Protocol (IP) address;
- a cookie ID*;
- the advertising identifier of your phone;
- data held by a hospital or doctor, which could be a symbol that uniquely identifies a person.

Persistent identifier: A persistent identifier is a long-lasting reference to a digital object that gives information about that object regardless of what happens to it. Developed to address “link rot,” a persistent identifier can be resolved to provide an appropriate representation of an object whether that object changes its online location or goes offline.¹⁵

Research data:

Research data are any information collected or generated for the purpose of analysis, in order to generate or validate scientific claims.

It includes digital and physical data. Research data encompass the whole spectrum ranging from raw data to the processed and analysed data. Examples include survey results, statistics, measurements, notebooks, images, texts, computer generated data, simulations, software developed for research purposes, computational metadata, prints, video- and audiotapes, coding of textual information, organisms, gene sequences, synthetic compounds, samples, patients data, etc.¹⁶.

Restricted access:

Open Access, but with restrictions. Often the Author wants to control the people who have access to for example a dataset. Access is provided in exchange for an e-mail address, to keep in touch with other people who are interested¹⁷.

¹⁴https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en#examples-of-data-not-considered-personal-data

¹⁵ <https://casrai.org/term/persistent-identifier/>

¹⁶ Research Data Management KU Leuven

¹⁷ <https://wiki.surfnet.nl/display/standards/info-eu-repo/#infoeu-repo-AccessRights>

Valorization: When your research is of interest for society in general or for a specific company, there is a possibility of valorization¹⁸.

Volumes: Volumes can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns¹⁹.

¹⁸ <https://bibliotheek.uhasselt.be/en/ethical-and-legal>

¹⁹ Science Europe RDM Practical guide, p.18