Data Management

This general factsheet presents the six main stages of the research data lifecycle. Each step is detailed in a dedicated factsheet. In addition, two cross-cutting sheets on the FAIR principles and legal, ethical, and scientific integrity aspects must be taken into account at every stage.

Before the project

1 Initial Planning

This first step involves the preliminary reflection and organization of the research project. It concerns the implementation of a Data Management Plan (DMP).

During the project

2 Description of Collected/Created Data

This step deals with the origin of the data, that is, the data collection methodology and the description of the datasets, with a preference for systems using metadata standards.

During the project

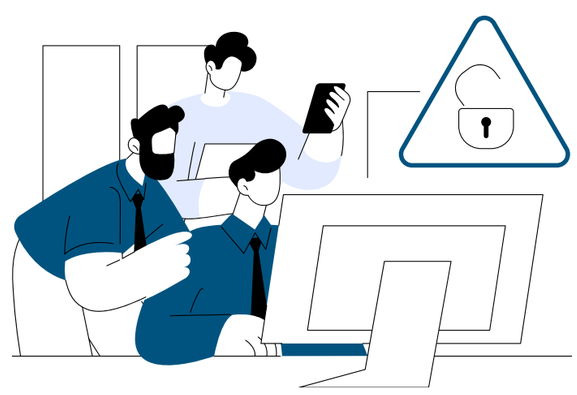
3 Validation, Processing, and Storage

A key stage of the research project, it includes data verification, cleaning, scientific validation, and processing. It also involves managing data through secure storage to ensure their integrity and accessibility.

After the project

4 Access and Sharing

This step involves publishing data papers, applying appropriate usage licenses to enable reuse by third parties, and selecting a suitable repository for depositing, accessing, and reusing data.



After the project

5 Long-term Preservation

This process aims to archive data for the long term, to maintain their intelligibility and accessibility over a period ranging from 10 to 30 years.

6 Reuse

This step supports the creation of new research, the re-examination of results, and data usage for teaching and learning, all while respecting the terms of the usage license.

CONTACT PROJET

Formulaire d'accompagnement LORD

Lille Open Research Data

SOURCE

DoRANum. Enjeux et Bénéfices : Le cycle de vie des données de recherche.
Disponible : https://doranum.fr/enjeux-benefices/le-cycle-de-vie-des-donnees-de-recherche_10_13143_gzj2-j593/

Auteur: Arthur Hemdat
Avec les conseils avisés de: Mónica Michel-rodrigez

1 Initial Data **Planning** and **Management**

LEM
Lille Economics Management

The Data Management Plan (DMP) is a key tool to support research data management and track the various stages of their lifecycle. Required by many funding agencies such as the ANR or Horizon Europe, it allows researchers to meet data management and sharing requirements.
Structured into different sections, the DMP should be updated regularly throughout the project. It helps to describe and monitor the evolution of datasets, while preparing for their sharing, reuse, and long-term preservation.

ADMINISTRATIVE INFORMATION

Includes project name, identifier, description, funding agencies, principal investigator, data contact person, date of first version and last update, as well as applicable policies.

DATA DESCRIPTION

Details the type, format, and volume of data, the datasets used, collection and creation methods, file organization systems, and quality assurance processes.

DOCUMENTATION AND METADATA

Covers the information needed to understand and interpret the data, how documentation and metadata were produced, and the adopted metadata standards.

STORAGE, BACKUP, AND SECURITY

Includes where data is stored, backup plans, those responsible for backup, recovery procedures, risk assessment, access control, and secure data transfer measures.

SELECTION AND PRESERVATION

Specifies which data will be kept, shared, or preserved, their intended use beyond the project, the archiving system chosen for long-term storage, and preparation steps.

DATA SHARING

Addresses how data will be made discoverable, any restrictions on data sharing, licensing terms, sharing mechanisms, publication timelines, and persistent identifiers.

LEGAL AND ETHICAL ASPECTS

Covers agreements for storing and sharing personal data, identity protection, sensitive data security, intellectual property rights, data ownership, reuse licenses, and usage restrictions.

RESPONSIBILITIES AND RESOURCES

Lists the person in charge of implementing the DMP, those responsible for data-related tasks, required tools and software, and any training or expertise needs.

LEGAL OBLIGATION

In France, Decree no. 2021-1572 of December 3, 2021 (Article 6), concerning compliance with scientific integrity, requires all public institution researchers to write a DMP.

ONLINE DMP TOOL

The tool DMP OPIDoR helps write your DMP online. It is freely available to the French higher education and research community.

TIMELINE

The DMP is a living document. Updates and specific deliverables may be required depending on the funders or project structure.

DID YOU KNOW?

A DMP can be written for both open data and **restricted** or **closed** access data. In the latter case, the plan must explain the reasons for non-disclosure.

OPEN ACCESS-ORIENTED

The DMP is closely tied to the open access principle. Depending on your choices and constraints regarding data sharing, specific criteria must be defined.

CONTACT PROJET

Formulaire d'accompagnement LORD ; Formulaire d'accompagnement du laboratoire

Lille Open Research Data

SOURCE

Inist-CNRS. (2017). DoRANum-Plan de gestion de données : fiche synthétique. DoRANum. https://doi.org/10.13143/CGV4-0K53

Auteur: Arthur Hemdat
Avec les conseils avisés de: Mónica Michel-rodrigez

2 Description of Collected/Created Data

In your Data Management Plan (DMP), it is essential to specify the origin of your data, whether they are collected, created, or reused.
For each data type, it is important to describe the collection methodology and provide a detailed description, in order to ensure rigorous data management aligned with the FAIR principles, and to facilitate understanding, sharing, and reuse of the data by other researchers over the long term.
This factsheet outlines the key elements to include when describing data in a DMP.

DATA DESCRIPTION

Every DMP must include a detailed description of the research data used in the project. Data should be described as precisely as possible to ensure they are understandable and usable.

Provenance
Are the data observational, experimental, computational, derived, or compiled?

Type
Are they textual data, numerical data, audiovisual data, models, scripts or codes, or specific data tied to a discipline?

Stability and Criticality
Are the data fixed, growing, or revisable? Are the data considered sensitive?

PROVENANCE

Administrative data
Collected during routine institutional activities (e.g., civil status registration)

Observational data
Captured in real time (e.g., survey data)

Experimental data
Obtained in a laboratory, reproducible but costly (e.g., experimental sessions)

Derived/compiled data
Resulting from processing raw data (e.g., indicators, econometric models)

TYPE

Textual data
Field notes, survey responses

Numerical data
Tables, measurements...

Audiovisual data
Images, videos

Specific data
Linked to a particular discipline or instrument

STABILITY AND CRITICALITY

Fixed
Data that cannot be altered once collected

Growing
New data added without modifying previous ones

Revisable
New data added with the possibility of modifying previous ones

Criticality
Are the data sensitive or confidential? (see the factsheet Legal, Ethical, and Scientific Integrity Aspects)

CONTEXT OF PRODUCTION

Collection
Are you collecting existing data? If so, provide the access path, source, and access conditions.

Creation
For newly created or generated data, describe the creation process or operational method used.

Reuse
Do you intend to reuse pre-existing datasets?

DATA FORMATS

To anticipate long-term archiving, the use of open formats is crucial to ensure long-term readability.

Open formats
Non-proprietary files, transparently encoded, part of the public domain. Ensure data accessibility and sustainability (e.g., CSV, TXT, PDF/A).

Closed formats
Fichiers propriétaires, les formats fermés n'appartiennent pas au domaine public. Ils requièrent l'utilisation de logiciels adéquats pour leur lecture et modification. (ex : DOCX, XLSX)

USE OF METADATA STANDARDS

Metadata provide information describing a dataset's essential characteristics (content, provenance, format, producer, etc.).

This is a recognized, standardized, and widely used method.

Example standards
Dublin Core : used to describe various types of digital resources.

METS : used to encode metadata in XML documents.

USE OF A README FILE

A README file may be needed to describe in more detail the production context and/or the data within the files.
It complements the metadata filled during dataset deposition, the data dictionary (which it may also include), and other available documentation.

It is usually disseminated in an open, widely used format, such as plain text (.txt) or markdown (.md).

CONTACT PROJET
Formulaire d'accompagnement LORD ; Formulaire d'accompagnement du laboratoire

SOURCE
DoRANum. Données de la recherche : apprentissage numérique. Plan de gestion de données : Origine et description des données de la recherche. Disponible sur : https://doranum.fr/plan-gestion-donnees-dmp/origine-description-donnees-recherche_10_13143_e9zh-w908/

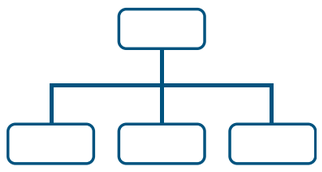
Auteur: Arthur Hemdat
Avec les conseils avisés de: Mónica Michel-rodrigez

3 Validation, Processing and Storage of Data

This factsheet covers the steps related to validation, processing, and storage of data within a Data Management Plan (DMP). Validation ensures data accuracy and reliability. Processing transforms data into usable information. Finally, secure data storage guarantees their accessibility and safety, in line with best practices and existing standards.

File Naming

- Give files a short and explicit name
- Do not use spaces or special characters
- Use dates in ISO 8601 format: YYYY-MM-DD
- Place the key element first for quick identification
- Indicate the file version (e.g., PV: provisional version, FV: final version...)



File Tree Structure

- Adopt an organization by theme (or topic)
- Structure folders hierarchically
- Limit the hierarchy to 4 or 5 levels
- Avoid folders named “Miscellaneous”
- Use clear and intelligible folder names

Data Preprocessing

Check for errors such as typos, inconsistent values, missing information, duplicates, imprecise entries, etc. This is referred to as preprocessing.

Data Storage

At LEM, the recommended institutional storage solution for data is s-Drive, the official CNRS service.

- Accessible à tous les membres du LEM via leur compte Janus (chaque agent CNRS/UMR en dispose automatiquement).
- Smooth synchronization via a web browser or NextCloud (similar to Dropbox/Google Drive).
- Personal storage space of 100 GB, non-expandable.
- Built-in recycle bin: deleted files are kept for up to 30 days.
- Secure hosting in France, with antivirus scanning and weekly backups.
- Option to create shared spaces (workspaces) for collaborative projects, with a dedicated quota.

Data Processing and Analysis

Analysis can be performed using tools such as visualization or machine learning.
Tracking modifications is crucial, especially for sensitive data.
A collaborative repository (forge) allows version tracking while facilitating collaboration.
Using a sovereign forge is recommended — for example, the University of Lille provides an institutional GitLab forge via its ENT.

CONTACT PROJET

Formulaire d’accompagnement LORD ; Formulaire d’accompagnement du laboratoire

SOURCE

DoRANum. Stocker ses données de façon sécurisée - Stockage et archivage.
Disponible sur : https://doranum.fr/stockage-archivage/stockage-donnees_10_13143_z0ge-nc29/

Auteur: Arthur Hemdat
Avec les conseils avisés de: Mónica Michel-rodrigez

4 Access and Dissemination of Data

LEM
Lille Economics Management

This factsheet presents the key means to ensure effective access and dissemination of research data within a Data Management Plan (DMP). This stage is based on three main levers: publishing data papers to increase the visibility and scientific impact of datasets, applying an appropriate usage license to govern third-party use, and selecting a reliable data repository to ensure their storage, accessibility, and reuse in an open science framework.

WHY PUBLISH A DATA PAPER?

- To raise awareness about the existence of the data and make them findable
- To credit the authors (recognition, citable reference) and valorize the dataset
- To facilitate data reuse and collaborative work (by making them intelligible)

DATA PAPERS

A data paper is a publication that describes a scientific dataset using structured information, known as metadata. Unlike traditional research articles, which test hypotheses or present new analyses, a data paper formalizes data sharing and follows a peer review process.

It is possible to generate a draft data paper from a dataset deposited in the Recherche Data Gouv repository.Click here for more information.

WHERE TO PUBLISH A DATA PAPER?

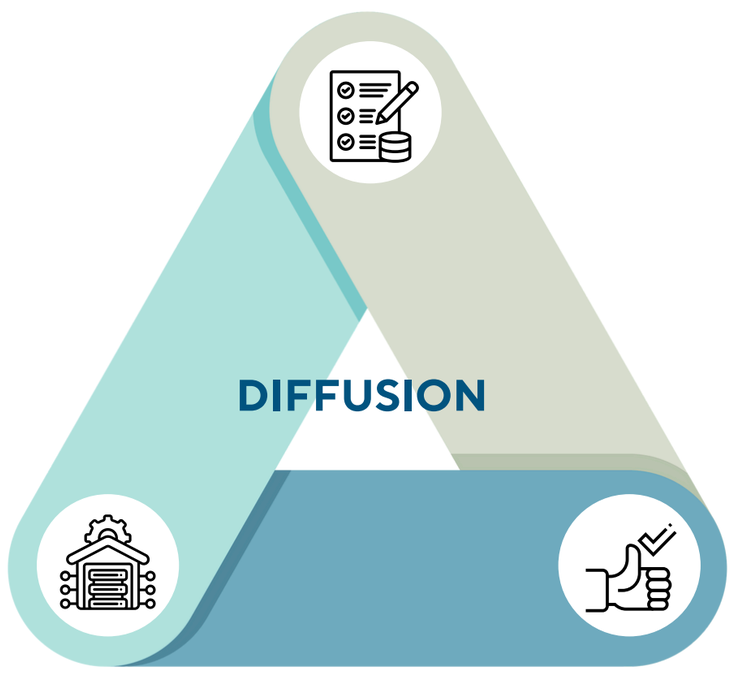
In a data journal, a journal dedicated to this type of publication
In a traditional journal that publishes data papers alongside research articles

WHAT DOES A DATA PAPER INCLUDE?

Data papers vary in structure depending on the journal, but they generally share common components

WHICH REPOSITORY?

Research data must be deposited on the recherche.data.gouv.fr platform, in the collection space of the Lille Économie Management (LEM) laboratory, via the University of Lille data repository, Lillodata.



REPOSITORY

To deposit scientifically validated research data (short to mid-term – 5 to 10 years) and make them accessible and reusable.
Several types exist: disciplinary, multidisciplinary, publisher-based, institutional, or project-specific.

LICENSE OF USE

A license is a contract whereby the data producer(s) authorize third parties to use the data and specify the intended purposes under defined access and reuse conditions. To promote data reuse, open licenses should be prioritized.

BEFORE DEPOSITING YOUR DATA

- The datasets to be shared have been selected
- Ethical principles are respected
- Rights of dissemination are verified
- Access conditions are defined
- Files are clearly named and organized
- Files are in sustainable and open formats
- The data are described and documented
- A license has been assigned to the dataset

WHICH LICENSES ?

To regulate the reuse of research data, you can choose from among the licenses available on data.gouv.fr, such as the Etalab Open License V2, which allows free and open reuse of data as long as the source is credited.

CONTACT PROJET

Formulaire d'accompagnement LORD ; Formulaire d'accompagnement du laboratoire

SOURCE

DoRANum. Enjeux et Bénéfices : Le cycle de vie des données de recherche.
Disponible : https://doranum.fr/enjeux-benefices/le-cycle-de-vie-des-donnees-de-recherche_10_13143_gzj2-j593/

Auteur: Arthur Hemdat
Avec les conseils avisés de: Mónica Michel-rodrigez

5 Data **Preservation**

Unlike data storage, the preservation phase — also known as long-term archiving — primarily aims to ensure the long-term conservation and future intelligibility of data.
Whereas storage focuses on data accessibility during the project, long-term preservation ensures durable retention (over 30 years) and involves financial, environmental, and scientific considerations.
Although important, this step is not mandatory, comes with a cost, and requires careful consideration of its relevance to the project's objectives.

PREPARING THE DATA

- Organize the data: define folder structure, naming, versioning, and appropriate file formats.
- Any document is either meant to be preserved permanently or destroyed once its administrative usefulness expires.
- Document the data using metadata standards to ensure traceability and future reusability.
- Estimate the data volume to anticipate storage needs.
- Schedule regular backups to avoid loss.
- Plan the necessary budget for data management and storage based on volume.

WHAT ARE THE PREREQUISITES?

Long-term preservation relies on three key principles to ensure data durability:

- Storage media integrity: ensure the long-term durability and security of the physical media.
- File format readability: use sustainable formats to avoid technological obsolescence.
- Data intelligibility: maintain the comprehensibility and usability of data through appropriate documentation (metadata, user guides, etc.).

KEYS TO LONG-TERM ARCHIVING

Research data preservation depends on dedicated digital archiving services (SAE).
In France, the main actor is the CINES, whose mission is to archive digital data and documents produced by the academic and research community.
CINES offers paid digital archiving solutions, for medium- and long-term durations, and provides expertise in both IT and archival science.
Data security and integrity are ensured through various procedures: metadata assignment, use of sustainable file formats, data replication, and a secure computing environment.
Use the services offered by CINES or other certified dgital archiving platforms to ensure proper preservation and compliance with best practices.

CONTACT PROJET

Formulaire d'accompagnement LORD ; Formulaire d'accompagnement du laboratoire

Lille Open Research Data

SOURCE

DoRANum. Données de la recherche : apprentissage numérique.Sockage et archivage : La minute stockage et archivage pérenne. Disponible sur : https://doranum.fr/stockage-archivage/la-minute-archivage-perenne-des-donnees-de-recherche_10_13143_77vj-0011/

Auteur: Arthur Hemdat
Avec les conseils avisés de: Mónica Michel-rodrigez

Les principes FAIR

In your Data Management Plan (DMP), it is essential to keep the FAIR principles in mind, as they apply to every stage of the research data lifecycle, starting from the very beginning of the project.
The FAIR principles (Findable, Accessible, Interoperable, Reusable) are international recommendations aimed at improving the management and sharing of research data.
They help make data easier to find, accessible, interoperable across systems, and reusable by other researchers — thereby promoting transparency and scientific collaboration.

FACILITATING DATA DISCOVERY

- Data must have a Persistent Identifier (PID) (e.g., a Digital Object Identifier or DOI) to ensure stable access.
- Data must be described using both scientific and documentary metadata.
- Data — or at least their metadata — must be indexed or registered in a searchable tool, such as a data repository or catalogue.

ENABLING **ACCESS** TO DATA AND METADATA

- Data must be accessible via the Internet, using a standard, open, and free protocol (e.g., HTTPS).
- Access may require authentication for restricted data.
- Metadata must remain accessible even if the data are temporarily unavailable or restricted.

FAIR PRINCIPLES

MAKING DATA **INTEROPERABLE**

- Data should be described from the start using controlled vocabularies.
- Metadata should, as much as possible, reference other datasets, allowing links between them.
- File formats should be open and documented to allow exploitation and long-term preservation using different tools.

ENABLING DATA **REUSE**

- Metadata must contain multiple relevant attributes to facilitate understanding and reuse.
- A reuse license must be assigned to the data.
- Data description must indicate provenance.
- The data structure must follow the standards of the scientific community to facilitate analysis.

CONTACT PROJET

Formulaire d'accompagnement LORD ; Formulaire d'accompagnement du laboratoire

Lille Open Research Data

SOURCE

Science ouverte - données de recherche", Collection Passeport pour la science ouverte (2024). Ministère de l'Enseignement supérieure et la recherche, Comité pour la science ouverte, Université de Lille.

Auteur: Arthur Hemdat
Avec les conseils avisés de: Mónica Michel-rodrigez

Legal, Ethical, and Scientific Integrity Aspects

In your Data Management Plan (DMP), this factsheet must be considered at every stage of the research data lifecycle. It supports the entire process, from data collection to sharing, by integrating key principles such as the protection of sensitive data, data sharing, and the rights and responsibilities of researchers.
Following these recommendations ensures rigorous and compliant data management, which is essential for ethical and responsible research.

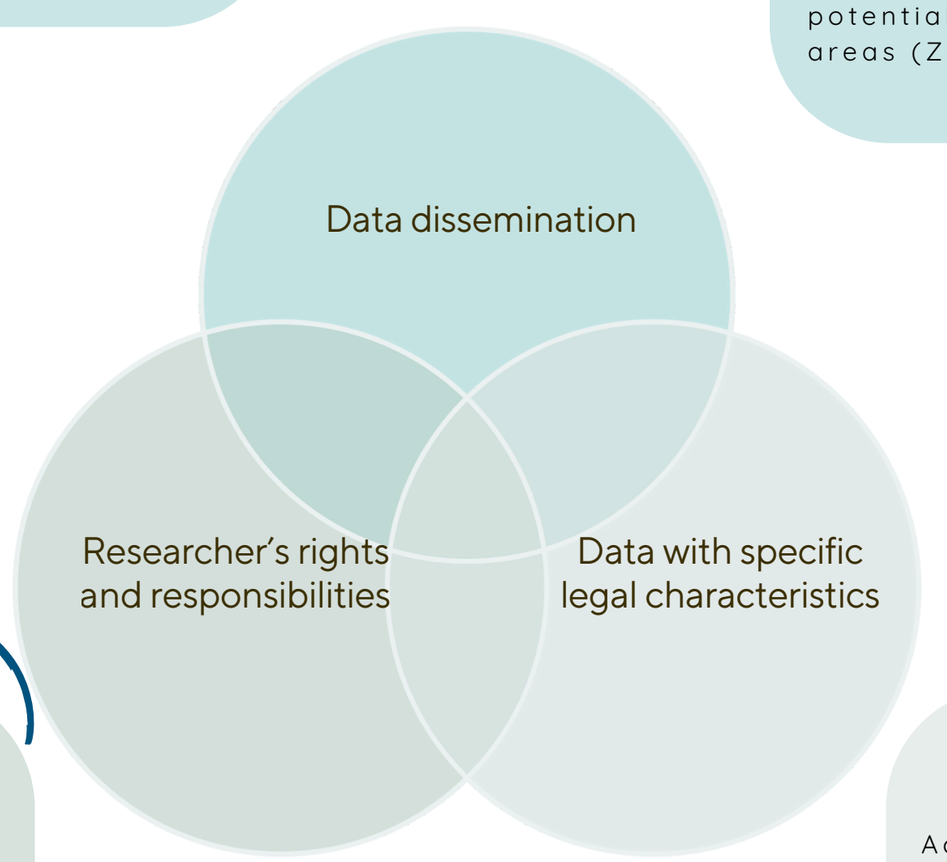
LEGAL FRAMEWORK FOR DATA

Since the 2016 Digital Republic Law, research data considered “finalized” and funded at least 50% by public funds are regarded as administrative data and are thus subject to a “default openness” principle.
Therefore, they are expected to be published and made accessible online, except in specific cases (e.g., personal data, sensitive data, industrial secrets, defense secrecy, scientific and technical potential protection (PPST), restricted areas (ZRR), etc.).

LICENSING

The license chosen by the author defines what the user is authorized to do with the data.
At a minimum, the user must respect the integrity of the data and acknowledge the authorship (i.e., source and last update date).

IESEG being a private institution, researchers affiliated with this organization are not subject to the same data openness obligations as researchers from public institutions.
Click here for more information.



APPLICABLE LEGAL REGIMES

Data are subject to database law. In this case, intellectual property rights also legally belong to the researcher’s home institution (e.g., CNRS, University of Lille, ULCO, IESEG, or University of Artois), which is considered the effective rights holder.

WHAT TYPES OF DATA ARE CONCERNED?

According to the CNIL, all types of data may be concerned: textual, numerical, audiovisual, specific...

ETHICS AND SCIENTIFIC INTEGRITY

Respect for privacy, intellectual property, data quality, and integrity are ethical dimensions of data management.
The European Code of Conduct for Research Integrity defines four core principles: reliability, honesty, respect, and accountability.

If you collect or process personal data under the GDPR, make sure to consult the Data Protection Officer (DPO) of your affiliated institution:

- DPO de l’Université de Lille
- DPO de l’IESEG
- DPO du CNRS
- DPO de l’Université ULCO
- DPO de l’Université d’Artois

WHAT ARE PERSONAL DATA?

Personal data are any information that can identify a person, directly or indirectly. This includes:
first and last name, age, gender, phone number, IP address, mailing and email addresses, voice, image, signature, fingerprints, etc.

WHAT ARE SENSITIVE DATA?

Sensitive data are a subset of personal data and may reveal:
racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, as well as the processing of genetic data, biometric data for unique identification, health data, and data relating to sexual life or sexual orientation.

CONTACT PROJET

Formulaire d’accompagnement LORD ; Formulaire d’accompagnement du laboratoire

Lille Open Research Data

SOURCE

Inist-CNRS. (2017). DoRANum-Aspects juridiques, éthiques, intégrité scientifique : fiche synthétique . DoRANum. https://doi.org/10.13143/WZ7A-HH81

Auteur: Arthur Hemdat
Avec les conseils avisés de: Mónica Michel-rodrigez