

Corinne Brusque
Céline Rousselot

How to create a data set in view of publication on the web in a data repository ?



Université
Gustave Eiffel

19 septembre 2025

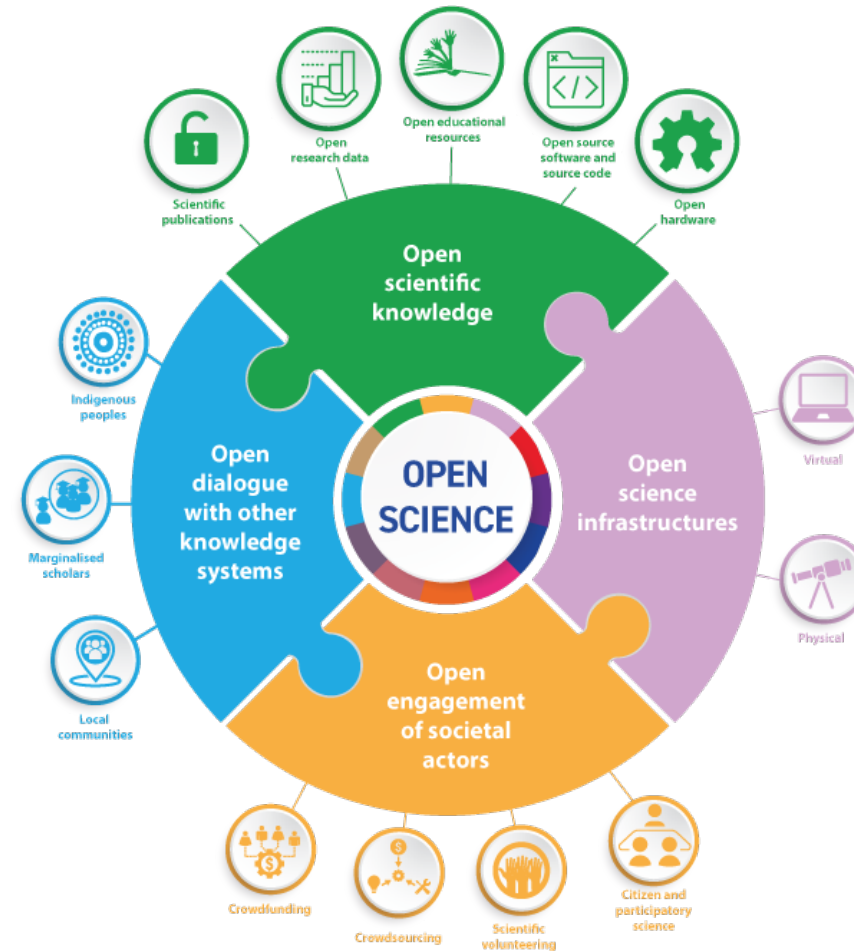
Definitions (1/3)

Open science

Open Science, is the movement to make scientific products and processes accessible to and reusable by everyone, as early and widely as possible.

Open Science encourages collaboration across academia, industry, public authorities, and citizen groups in order to increase creativity and trust in science.

This engagement process leads to greater transparency in the research process, greater potential impact of research, more efficient research processes and opportunities for global scientific collaboration.



UNESCO. (2021)

Definitions (2/3)

Research data

Research data are defined by OECD:

- as factual records in the form of figures, texts, images and sounds
- which are used as the main sources for scientific research
- which the scientific community generally recognises as being necessary to validate research results.

Definitions (3/3)

Open research data

Open research data is data that can be freely accessed, reused, remixed and redistributed, for academic research and teaching purposes and beyond.

Ideally, open data have no restrictions on reuse or redistribution, and are appropriately licensed as such.

In exceptional cases, e.g. protection of the identity of human subjects / protection of property rights, restrictions of access are set.

Openly sharing data exposes it to inspection, forming the basis for research verification and reproducibility, and opens up a pathway to wider collaboration.

Open Science in Horizon Europe 2021-2027

Two mandatory practices:

Open access to publications and to research data based on the principle of 'as open as possible, as closed as necessary' and on the FAIR principles.

What does "as open as possible, as closed as necessary" mean?

Results and data may be kept closed if making them public in open access is against the researcher's legitimate interests.

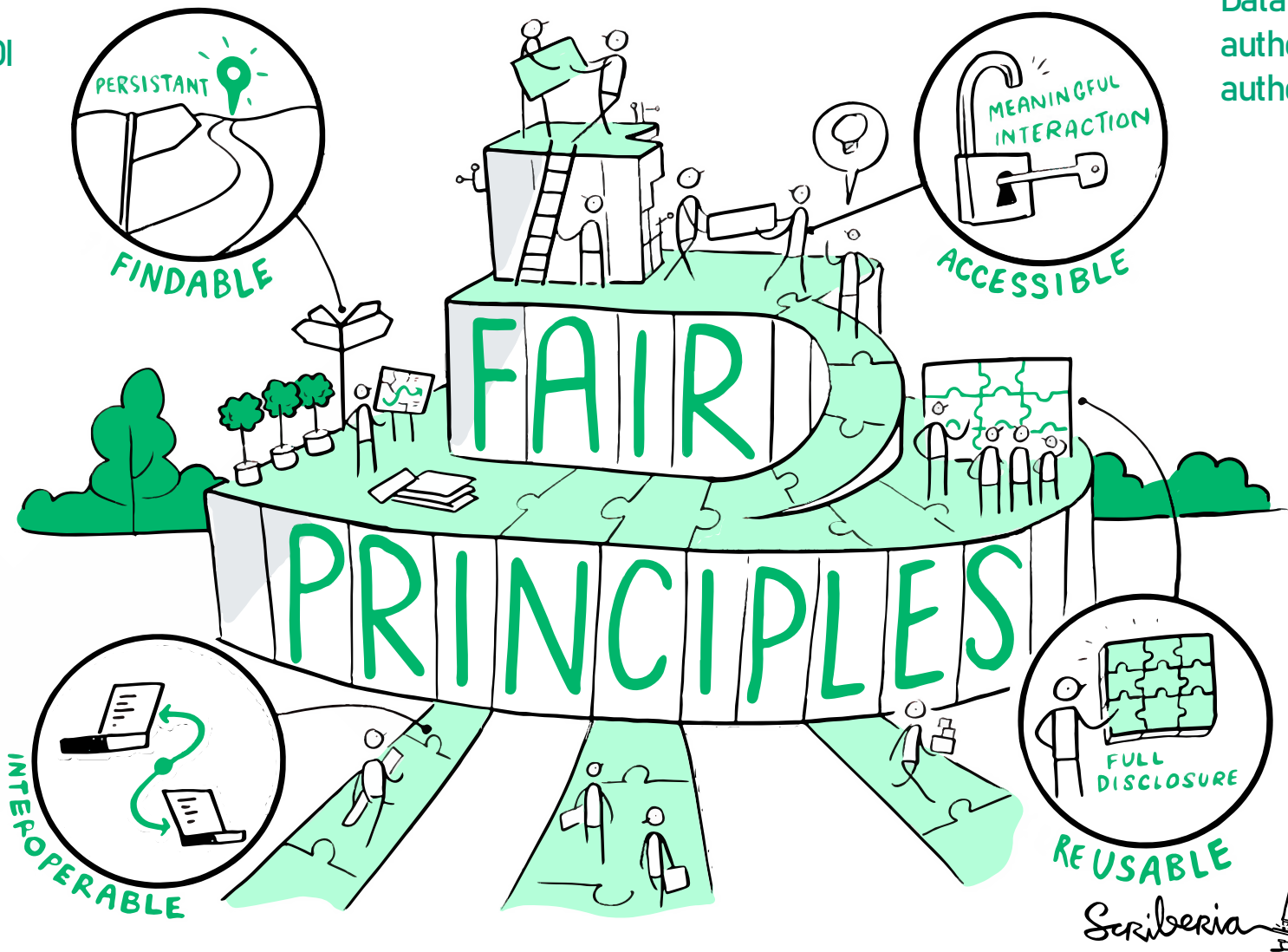
Examples include to commercially exploit their research results, or if it is against any obligations mentioned in the Grant Agreement (e.g. personal data protection).

What are the FAIR principles?

The research data of EU-funded projects must be FAIR (Findable, Accessible, Interoperable and Re-usable).

The FAIR principles emphasise the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention.

Descriptive data
Persistent identifier, DOI



Data openly available
authentication and
authorisation procedures

Metadata standards
Controlled vocabulary

Documentation
Licence

Selection of a trusted data repository (1/2)

A data repository is an online archive where researchers deposit and share research data.

Example of criteria used to identify the quality of a data:

- Type of moderation
- Assignment of a persistent identifier
- Sustainability of the infrastructure
- Possibility of embargo
- Type of data
- Volume limit
- Policy on transfer of rights
- Pricing policy
- Location of data outside the European Union

Frédéric de Lamotte, Véronique Stoll, Cécile Arènes, Marie-Emilia Herbet, Stéphane Debard, et al.. Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante. Comité pour la Science Ouverte. 2024. (hal-04534321v2)

Data Repository Certification

- <https://amt.coretrustseal.org/certificates>
- [Nestor Seal](#): verification according to DIN 31644
- [ISO 16363 certification](#)



Selection of a trusted data repository (2/2)

A data repository is an online archive where researchers deposit and share research data.

The general steps for finding a data repository are:

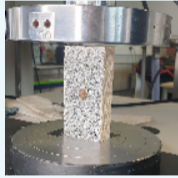
1. Explore the available data repositories
 - the Directory of Open Access Repositories, <https://opendoar.ac.uk/>
 - Re3Data, <https://www.re3data.org/>
 - FAIRsharing, <https://fairsharing.org/>
1. Use a thematic/disciplinary repository that would align with your research.
 - Open Research Europe-approved repositories for physical sciences and engineering : <https://open-research-europe.ec.europa.eu/for-authors/open-data-software-and-code-guidelines-pse/#approvedrepositories>
2. Alternatively, use an institutional repository, if you have access to one which is a trusted one.
 - <https://entrepot.recherche.data.gouv.fr/dataverse/univ-gustave-eiffel>
3. Should you be unable to identify a suitable repository, Zenodo is a centralized repository that is supported by the European Commission and helps beneficiaries in meeting their Open Access requirements.
 - <https://zenodo.org/>

Some examples of datasets on <https://entrepot.recherche.data.gouv.fr/>

Recherche Data Gouv > Data Univ. Gustave Eiffel >

Material parameters of a phase-field model used to simulate brittle fracture of spruce specimens

Incomplete metadata Version 6.1



Noel, Matthieu; Pled, Florent; Chevalier, Luc, 2024, "Material parameters of a phase-field model used to simulate brittle fracture of spruce specimens", <https://doi.org/10.57745/NGOKFP>, Recherche Data Gouv, V6

Cite Dataset ▾

Learn about [Data Citation Standards](#).

Access Dataset ▾
Contact Owner Share

Make Data Count (MDC) Metrics ⓘ

since 2020-07-01

1,466 Views ⓘ

89 Downloads ⓘ

Description ⓘ

Data used to identify the material parameters of a phase-field model for brittle fracture of spruce specimens. [Link to the SpruceParams GitLab repository](#) [Recherche Data Gouv > Data Univ. Gustave Eiffel >](#)

Subject ⓘ

Engineering; Physics

Keyword ⓘ

Brittle Fracture, Cracking, Phase-field, Transversely isotropic testing, Cracks

Related Publication ⓘ

Noel, Matthieu, Pled, Florent, Chevalier, Luc, Wilquin, François phase-field modeling of brittle fracture in spruce wood. *Engine* 2025. DOI: 10.1016/i.endfracmech.2025.111304. HAL: hal-05

Challenges of embedding fiber Bragg grating sensors in castable material: influence of the material shrinkage and fibers coatings on ultrasonic measurements

Incomplete metadata Version 1.1



DERRIEN, Nicolas, 2025, "Challenges of embedding fiber Bragg grating sensors in castable material: influence of the material shrinkage and fibers coatings on ultrasonic measurements", <https://doi.org/10.57745/9IYNUO>, Recherche Data Gouv, V1

Cite Dataset ▾

Learn about [Data Citation Standards](#).

Access Dataset ▾
Contact Owner Share

Make Data Count (MDC) Metrics ⓘ

since 2020-07-01

470 Views ⓘ

19 Downloads ⓘ

0 Citations ⓘ

Description ⓘ

This dataset is linked to a paper on the feasibility of ultrasonic measurements with Fiber Bragg Grating (FBG) embedded in polyurethane resin samples. It combines data on the effect of resin on the reflectivity spectra of FBGs and ultrasonic measurements made with these same sensors. Initially, this measurement campaign was designed to determine which of several types of optical fiber (polyimide coating, acrylate coating, bend-insensitive, 7-core) was most suitable for insertion into polyurethane resin. English (2025-02-19)

- <https://doi.org/10.57745/NGOKFP>
- <https://doi.org/10.57745/9IYNUO>

Before depositing (1/3)

Making sure the data can be shared

by checking:

- that any co-authors agree
- contracts or regulations
- if there is any sensitive data
- Personal data :
Anonymization , Agreement
of persons

Examples of sensitive data are:

- **Personal data:** identifiers such as names or identification numbers, physical, physiological, genetic, mental, economic, cultural or social characteristics, it also includes location data from GPS or mobile phones
- **Confidential data:** trade secrets, investigations, data protected by intellectual property rights Security: passwords, financial information, national safety, military information...
- **Combination of different datasets** that can be combined into sensitive or personal data
- **Biological data:** endangered (plant or animal) species, where their survival is dependent on the protection of their location data (biodiversity community) ↗
- **Personal and sensitive metadata** ↗

Before depositing (2/3)

Preparing the data files

Make sure the files comply with the following best practices and also any recommendations within your discipline if such exist

- naming conventions,
- an open or widely used community file format which is compatible with long-term preservation
- the coherency and organisation of the data within a dataset

Before depositing (3/3)

Preparing the documentation of the data

Please provide all the information necessary to understand your data.


- Some information will be provided through the metadata: publications, data management plan, the project number and funding, keywords, datasets or other research products linked to the deposit.
- Each file or set of files must be accompanied by information about the data that enables it to be reused. This can be in different forms:
 - A self-documenting data format (NetCDF - Network Common Data Form)
 - A data dictionary that lists and defines the metadata of data contained in a database, file or application.
 - A README file which provides complementary information. Read me template : <https://recherche.data.gouv.fr/en/category/33/guide/readme-template>

Source : <https://recherche.data.gouv.fr/en/online-help>

The ins and outs of metadata and data documentation

Metadata and data documentation

Copier le li...



Metadata

- Species: Horse
- Distance: 3.2
- Outside temp: 24.3
- Owner: Category 1
- Description: Pr02_urban_2017.txt
- Project: TempDiagn2017

Picture CC BY-NC 4.0
From: Rushton JD, Tichy A, Neil B.
Introduction of the use of thermography and thermometry in the diagnosis of ulcers in horses: a pilot project.
VetRecOpen 2015; 2 :e000089. Doi:10.1136/vetreco-2014-000089

Plein écran

6:13 / 6:13

YouTube

Source : <https://www.uu.nl/en/research/research-data-management/guides/during-research/metadata-and-documentation>

Despositing data

- Building up of the data set
 - Creating the dataset
 - Adding the associated files to a dataset
 - Completing the metadata of the dataset
- Indicate the terms of use for the datas
 - Choice of the Licences (different for dataset and software / permissive or not)
 - Conditions for restricted access files
- Managing the rights associated to datasets and files
- Scientific publications linked to the data
 - Applying an embargo to a data file
 - Giving access to an unpublished dataset (private URL)
 - Giving access to an unpublished dataset (URL for anonymized access)

- <https://ukdataservice.ac.uk/learning-hub/research-data-management/rights-in-data/licensing/>

- <https://interoperable-europe.ec.europa.eu/collection/eupl/solution/licensing-assistant/find-and-compare-software-licenses>

Publishing data

- The dataset becomes visible on the web.
 - The metadata is public.
 - Data files are in open access unless the depositor applies restrictions.
 - A permanent identifier, such as DOI, is assigned to the dataset.
- The published dataset cannot be deleted.
- Additional versions of the dataset can be published.
- The dataset can be cited and a citation format is provided.
- The dataset can be linked to publications.

Thank for your attention

