

Open Science Guide of the EDSP2

Principles and Definition of Open Science

Table of Contents

Principles and Definition of Open Science

Why Open Science?

The advantages of open science for a doctoral student

How to open science?

	3
A) Generalize open access to publications	3
1- Deposit your results in an open archive	5
2- How to know the journal's policy	
3- Predatory journals	5
4- Open access rights	5
5- ORCID identification	7
6- Open U Journals: a Bordeaux solution to create a journal	8
7- The HAL institutional portal of the University of Bordeaux.....	8
B) Structure and open research data	8
1- Prepare and document data	9
2- How to choose a data repository?	9
3- Data repository selection aid tool	9
4- ABDo: the Bordeaux workshop for research data	10
5- The institutional space Research Data Gouv of the University of Bordeaux	10
6- Link data to publications	11
7- FAIR and DMP	11
C) Engage in a sustainable, European and international dynamic	13
To go further: the training offer in ADUM and the sources	14

Principles and definition of Open Science

Open science aims to make knowledge, data, methods, and research results accessible to all, in a transparent and collaborative way.

It is based on several principles:

- Open access to publications: scientific articles are freely available online, without subscription;
- Sharing of research data: data produced by researchers can be reused and verified;
- Transparency of methods and codes: protocols, software, and tools used are open and reproducible;
- Collaboration and broad dissemination: researchers, citizens, businesses and institutions can participate more easily in the production and use of knowledge.

The main objective is to make research :

- more accessible;
- more reproducible;
- more efficient;
- and more useful to society.

Why open science?

Three main factors should encourage broader practice of open science. They are scientific, economic, and civic.

- Scientific: science that shares the products of scientific research enables the circulation of protocols, data, and results. Interaction between researchers, reproducibility (and thus transparency and scientific integrity) and reinterpretation are fostered. Moreover, openness can also, thanks (among other things) to new data processing tools, open new research avenues and foster innovation.

- Economic: A more open science aims to propose an alternative model to the current economic model of scientific publishing. Indeed, the constant rise in costs is no longer sustainable even as digital tools today allow scientists to regain control of the publishing circuit.

- Citizen: A more open science responds to citizens' demand for transparency and helps combat false scientific information that circulates today in the highest levels of decision-making bodies.

The advantages of open science for a PhD student

- Increased visibility and impact Open access publications and data are more cited and disseminated. For a PhD student building their scientific reputation, it is an opportunity to publicize their work far beyond their discipline or institution, from the start of their career.

- **Strengthened scientific credibility and integrity** Sharing one's data, protocols, and source code demonstrates the rigor of their approach. This also protects the doctoral candidate in case of a challenge: everything is traceable and verifiable, which strengthens the community's trust in their results.
- **Acceleration of thesis progress** Freely accessing the data and publications of other researchers avoids "reinventing the wheel." The doctoral candidate can rely on existing resources, reproduce analyses, and save valuable time in their own research.
- **Opportunities for collaboration and networking** Making one's work accessible draws the attention of other researchers, potentially in complementary fields. This can lead to unexpected collaborations, co-publications, or invitations to conferences— all of which are advantages for an academic or industrial career.
- **Preparation for institutional requirements and funders** Increasingly, funding agencies (ANR, Horizon Europe...) and universities require depositing in open access as a condition for obtaining funding. Training in these practices from the doctoral stage means anticipating obligations that will be unavoidable throughout your career.

How to open science?

3 major lines of action:

- A) Generalize open access to publications
- B) Structure and open research data
- C) Take part in a sustainable, European and international dynamic

A) Generalize open access to publications

1- Deposit your results in an open repository

Three main routes make it possible to achieve this goal: depositing in an open archive (the green route), publishing in open access (the gold route), publishing in open access journals (the diamond route).

The green route

The green route involves, for researchers, depositing a version of their scientific publications in an open archive, regardless of whether they are published in a scientific journal.

- HAL (Hyper Articles en Ligne): is a multidisciplinary open archive developed in 2001 by the Centre for Direct Scientific Communication (CCSD), a unit of the CNRS.

- arXiv: created in 1991 by Cornell University, arXiv is considered one of the oldest open archives. It covers research areas in physics, mathematics, computer science, and quantitative biology.

- PubMed Central: launched in February 2000 by the National Library of Medicine of the National Institutes of Health (NLM/NIH) in the United States, PubMed Central provides access to the full text of articles in the biomedical and life sciences domain.

The golden road

The golden path corresponds to the publication of a scientific article in open access in a journal, making its content immediately freely accessible to all readers. The general principle is to transfer publication costs to a third party other than the reader or subscriber. Three main models of open access journals are distinguished and differ

mainly by their funding model and access to content.

- Full open access journals (or “free”) : These are peer-reviewed journals whose all articles are freely accessible from publication. Readers do not pay to access the content, and authors generally do not pay publication fees (examples: OpenEdition, epijournals or Open U Journals). The Directory of Open Access Journals (DOAJ) lists these journals based on quality criteria, editorial transparency, and respect for copyright.

- The author-pays model (APC): In this model, open access publication is funded by fees called Article Processing Charges (APC), paid by authors, their lab or their institution. The article is then immediately freely accessible to all readers.

- The hybrid model: These are journals mainly funded by subscriptions, but which offer authors to make their article open access for the payment of APC. This means that in the same journal some articles are only accessible by subscription and others are open access. This model is often criticized because it can lead to double payment: institutions pay both for subscriptions to read the journals and the APC to publish in open access. For this reason, many open science policies encourage limiting or avoiding this model. Publication in hybrid journals (subscription + APC) is no longer funded by ANR or Horizon Europe.

The diamond path

The diamond path differs from the green and golden routes in its economic model. Unlike the green path, the diamond path directly concerns publication in open access journals. It also differentiates itself from the classic golden path in that it does not rely on payment of article processing charges (APC) by the authors or their institutions.

This model is generally funded by public institutions, universities, libraries, research organizations or public grants. This funding directly supports diamond journals, i.e., fully open access scientific journals that do not charge reading or publication fees. These journals rely on the collaborative work of researchers, librarians, universities and public platforms for scientific publishing, in a logic of academic cooperation rather than commercial.

Diamond journals are distributed and accessible via different publication platforms, such as OpenEdition Journals, Episciences or certain journals hosted by universities and research infrastructures.

2- How to know the journal's policy

Several tools allow checking a publisher's open access policy before publishing a research work.

- SHERPA/RoMEO: this service developed by the University of Nottingham catalogs publishers' open access policies and classifies them according to different levels of dissemination and archiving permissions.
- Willo: service developed by the University of Lille, allowing to know the rights linking an author to their publisher regarding open access for a given article.
- Directory of Open Access Journals (DOAJ): DOAJ lists open access scientific journals according to several criteria: free full-text access, transparency of editorial policy, permanent archiving, use of persistent identifiers, respect for copyright and use of Creative Commons licenses.

DOAJ also allows access to article metadata, thereby improving the visibility of scientific content. Its advanced search mode helps find journals that are fully open access, with APC or hybrid.

3- Predatory journals

Predatory journals are scientific journals that prioritize commercial objectives over academic standards and scientific ethics.

Although predatory publishers existed before the rise of open access, their recent proliferation has been promoted by the emergence and success of the author-pays model. These publishers exploit the author-pays publication fee model to generate profits.

These journals publish articles without a rigorous peer review process, despite assurances displayed on their sites. They typically offer very rapid publication times, low publication fees, an absent or superficial evaluation process, and editorial boards that are not transparent or filled with poorly identifiable researchers. Their article acceptance rate is also very high.

Predatory journals frequently contact researchers by mass sending unsolicited emails (spam) and particularly target young researchers, especially after publishing their thesis.

4- Open access rights

The Law for a Digital Republic (2016)

Article 30 of Law No. 2016-1321 of October 7, 2016 for a Digital Republic (known as the “Lemaire Law”), codified at Article L. 533-4 of the Code of Research, creates a new right for authors: the right to freely disseminate their scientific writings in open access, regardless of the contract signed with the publisher. This right applies to any writing published in a journal that has appeared at least once a year, provided that the research was financed at least 50% by public funds (state grants, local authorities, public institutions, national funding agencies or European funds).

- Without embargo when the article is already accessible (there may be APC costs) on the journal’s site ;
- This right allows the deposit of the accepted version for publication (postprint/AAM), with a maximum embargo of 6 months (STM) or 12 months (SHS), without the publisher being able to oppose it.

The Plan S and the acceleration of open access

Drafted in September 2018 by cOALition S, a consortium of 16 national research funding agencies supported by the European Commission, Plan S aims to accelerate the dissemination of scientific publications in open access. It states, in particular: From 2021, all scientific publications arising from research funded by public or private grants awarded by national, regional and international funding bodies must be published in journals or on open access platforms, or be immediately available without embargo in open archives.

This plan includes 10 principles available at

<https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles>

Obligations of funders: ANR and Horizon Europe

Beyond the open rights established by the 2016 Digital Law, major research funders have made open access deposition mandatory within their funding terms.

- National Research Agency (ANR): since the 2022 call for proposals (AAP), all publications arising from ANR projects must be available in immediate open access under CC-BY. Authors must deposit the full text (publisher’s version or accepted manuscript) in the national HAL open archive, indicating the ANR project decision code (e.g., ANR-22-CE64-0001), to enable automatic publication tracking. The Non-Derivation of Rights (SNCD) must be applied at submission. Additionally, a Data Management Plan (DMP) is required within 6 months of project start, and a final version is expected at closure.

- Horizon Europe: the European Framework Programme for financing requires immediate open access (no embargo) to all peer-reviewed scientific publications, under an open license (CC-BY for articles). A copy of the publisher’s version or accepted manuscript must be deposited in a trusted repository upon publication. Data management in accordance with FAIR principles is mandatory, and a Data Management Plan (DMP) must be provided within the first 6 months of the project, and regularly updated. Costs related to open access (APCs for fully Open Access journals) are eligible for funding provided they have been budgeted in the proposal.

Which version to submit?

- Preprint version: This is the initial version of the article written by the author before peer review, or before submission to a journal, or after rejection of publication. This version can often be freely deposited in an open archive. The rules of each publisher can be checked on SHERPA/RoMEO.

- Postprint version (Author Accepted Manuscript – AAM): This is the final manuscript version accepted by the publisher after peer review, but without the journal’s formatting. The law for a Digital Republic authorizes its open access deposition, subject to respect for any disciplinary embargoes.

- Publisher version: this is the final version published with the publisher’s layout. The law does not authorize its deposition in an open archive, unless the publisher has explicitly authorized it in its policy.

Creative Commons licenses

Scientific work deposited in open archives is protected by copyright, in accordance with the Intellectual Property Code. Their reuse remains subject to the conditions defined by the author. It is therefore recommended to associate a Creative Commons license with the deposit to specify the diffusion and reuse conditions.

Creative Commons licenses are free licenses that define the uses allowed by the author for any work of the mind. The author can authorize copying, distribution, and, in some cases, modification of the work. Creative Commons particularly recommends the CC BY 4.0 license as the standard for open access.

The Rights Retention Strategy (RRS)

This strategy allows authors to legally disseminate their manuscript in an open archive like HAL, before and after peer review, regardless of the publication venue.

Recommended by the National Research Agency since the 2022 project calls, it involves informing the publisher, from the submission of the article, that an open CC BY license applies to the different versions of the manuscript up to the accepted version. It is also possible to associate the Etalab 2.0 license, recommended by the French state and legally equivalent to CC BY. This mechanism allows immediate open access deposition in HAL upon publication, with no embargo. This strategy is supported by France Universités and CNRS.

5- ORCID identification

ORCID is a non-profit organization that assigns persistent digital identifiers to researchers and scientific contributors worldwide.

The ORCID identifier helps avoid confusion related to homonyms among researchers. It also groups the different forms of identity under which an author may publish (name variations, initials, name changes, etc.).

It links together the scientific works produced by a researcher, even when they are spread across different databases or information systems. It also facilitates interoperability between platforms by avoiding repeated re-entry of personal information and bibliographic references.

ORCID can also be linked to IDHAL, the author identifier used in HAL. This connection facilitates communication between the two systems, enables synchronization of information, and prevents researchers from re-entering the same data multiple times.

6- Open U Journals: a Bordeaux solution to create a journal

Open U Journals is a digital platform offered by the University of Bordeaux since 2016 on a Diamond model.

The platform allows the creation and management of open access scientific journals. It offers researchers on the Bordeaux site, as well as other scientific teams, the ability to create and publish journals, ensuring the entire editorial process. This includes article submission, peer review, and online publication in open access.

A scientific committee reviews each journal creation request to verify its integration into the system. Journals must meet the eligibility criteria recommended by the Committee for Open Science.

7- The University of Bordeaux's institutional HAL portal

The University of Bordeaux has its own institutional HAL portal, accessible at <https://u-bordeaux.hal.science>.

This portal allows members of the university's scientific community to deposit, promote, and make their publications openly accessible. The university libraries offer personalized support for deposits (rights verification, quality control of records, HAL workshops) via the contact: info@oskar-bordeaux.fr for publications.

B) Structure and open research data

Making research data as accessible and reusable as possible addresses several major challenges.

- Promote innovation and collaboration: enable data reuse for potential new applications and encourage the creation of collaborations between data producers and users.
- Ensure data preservation and reduce costs: preserve data that is unique or costly to produce, avoid costs associated with duplicating existing data, and limit the risk of data loss.

- Ensure scientific transparency: facilitate reproduction and replication of research, enable verification and accuracy of results, and help deter scientific fraud.
- Value researchers' work: generate direct recognition for researchers who produced the data, notably through the publication of data papers¹ or by adapting citation practices. Some journals encourage not only citing datasets but also integrating them into the main bibliography of scientific articles.
- Meet institutional and regulatory obligations: apply the requirements of scientific journals or funding agencies, such as those of the European Commission.

1- Prepare and document data

Preparation and documentation of research data, including metadata, are essential to ensure their understanding and reuse. The Data Management Plan (DMP) helps organize this work throughout the project.

Data sharing must take into account legal constraints and institutional policies. It is recommended to avoid supplementary data proposed by publishers, often poorly suited for good documentation and effective reuse.

It is preferable to deposit data in institutional, disciplinary, or general repositories that ensure their longevity, accessibility, and reuse. Finally, the link between data and publications should be anticipated from the start of the project.

2- How to choose a data repository?

The recommended approach for sharing research data is to:

- Favor a disciplinary data repository when one exists, for example PANGAEA, 4TU.ResearchData, NAKALA, or Harvard Dataverse;
- If not, use a generalist repository such as Zenodo or B2SHARE;

The choice of a data repository is based on several criteria, including supported formats, indexing possibilities, and data access modalities (open access, restricted, or embargoed). Using certified repositories is also an important criterion to ensure quality, longevity, and reliability of the service offered. The re3data.org portal helps identify these repositories using its 'Certificates' filter.

3- Data repository selection aid tool

The Documentation Department of the University of Bordeaux provides a freely accessible online decision-support tool titled 'Find a Data Repository'. This

¹ A data paper is a scientific publication whose primary objective is to describe a research data set. Unlike a standard scientific article, it provides a detailed description of a data set and associated metadata, as well as collection conditions and methods, without presenting results based on its analysis.

the tool makes it possible to identify the most suitable general-purpose data repository for researchers' needs based on four main criteria:

- The need to restrict access to the data;
- The existence of free deposit options;
- The ability to freely define the conditions for using the data;
- Review of the deposit before it is published online.

A concise descriptive sheet accompanies each of the eight data repositories compared: Zenodo, Figshare, Mendeley Data, Dryad, B2SHARE, 4TU.ResearchData, Open Science Framework and Harvard Dataverse.

4- ABDo: the Bordeaux Workshop for Research Data

ABDo (Bordeaux Workshop for Research Data) is a local service dedicated to researchers, PhD students, engineers and research support staff from all institutions at the Bordeaux site. Labeled by the Ministry of Higher Education and Research, ABDo offers a single desk (address: donnees-recherche@u-bordeaux.fr) for: training and awareness-raising on research data management and the FAIR principles; support for depositing, storing, sharing and archiving data; legal and ethical advice (intellectual property, licenses, GDPR); assistance with drafting Data Management Plans (DMP/PGD); and networking for peer exchanges of best practices.

5- The institutional space Recherche Data Gouv of the University of Bordeaux

The University of Bordeaux has a dedicated institutional space within the national multidisciplinary repository Recherche Data Gouv, the French sovereign platform dedicated to sharing research data (<https://entrepot.recherche.data.gouv.fr/dataverse/univ-bordeaux>). Based on the open-source software Dataverse and developed in particular by INRAE, this repository is a preferred solution for depositing institutional data for units seeking a sovereign national solution. Accessible to all members of the university via their UBx identifier (SSO), it allows users to deposit or group datasets by organization, team or project. (contact: donnees-recherche@u-bordeaux.fr).

Whatever trusted repository is selected for data sharing, it must guarantee a minimum set of features ensuring the dissemination, traceability, preservation and reuse of research data. These features also help to comply with the FAIR2 principles.

A data repository must in particular make it possible to:

2 The FAIR principles are based on four dimensions: Findable (easy to find), Accessible (accessible), Interoperable (interoperable), Reusable (reusable).

- The assignment of a persistent identifier (Persistent Identifier – PID) of the DOI type, allowing citation, traceability, and interconnection of data with other scientific outputs, such as publications ;
- A sufficiently detailed description of the data to facilitate discovery, understanding, and reuse, through standardized metadata and the use of controlled disciplinary vocabularies ;
- The use of licenses and the definition of access rules to legally frame data reuse, in compliance with French and European law ;
- A minimum retention period for data aligned with the retention policy defined by the affiliated institution.

6- Link data to publications

Several solutions allow establishing a link between a scientific article and the data associated with it before publication. The use of persistent identifiers thus facilitates durable interconnection between publications and research data. Similarly, indexing publications associated with the data (including data papers) is generally possible in most data repositories, even after the initial deposit of datasets.

After publication of an article, adding a link to the associated data can be difficult, or even impossible, on some editorial platforms. An alternative is to record these data in the deposited version of the article in an open archive such as HAL, which allows adding persistent identifiers in the 'Associated data' field. This thus facilitates linking between publications and data sets.

7- FAIR and DMP

The FAIR Principles

The FAIR principles aim to promote discovery, access, interoperability, and reuse of research data. They define a set of characteristics that data and their metadata must have to facilitate use by humans as well as machines.

The INRAE website offers detailed resources in French on the FAIR principles:

<https://science-ouverte.inrae.fr/fr/le-numerique-pour-la-science-et-les-donnees-scientifiques>

10 actions to make your data more FAIR

collaboratively developed guides offer concrete recommendations to improve the FAIRification of data and software. The resource titled Top 10 FAIR Data & Software Things presents, for different disciplinary domains, ten key actions to make data and code easier to find, access, interoperate, and reuse. Each recommendation is accompanied by practical examples.

These guides are available in open access as PDF documents, notably via the Zenodo platform, under the following reference: Christopher Erdmann et al. (2019), Top 10 FAIR Data & Software Things.

How to evaluate the FAIR level of data?

Several tools allow evaluating to what extent data comply with the FAIR principles:

- How FAIR are your data?: a simple checklist in checkbox form enabling quick self-assessment ;
- FAIR Self-Assessment Tool: an online tool developed by the Australian Research Data Commons, which generates an overall score as well as detailed scores for each FAIR principle.

The latter offers a series of questions for each principle, with answers in the form of dropdown menus, enabling structured and reproducible evaluation.

The DMP

The DMP (Data Management Plan), or PGD (Plan de Gestion de Données), is a concise document that helps organize and anticipate the entire lifecycle of research data sets. It describes, for each data set, the management modalities implemented from their creation or collection to their sharing and archiving.

According to INIST-CNRS, it is a document “that helps to organize and anticipate all stages of the data lifecycle” and that specifies “how the data of a project will be managed, from their creation or collection to their sharing and archiving” (INIST-CNRS, 2018, Data management plan: summary sheet, DoRANum).

The content of a DMP varies depending on the model used, whether imposed by an institution or chosen freely. It generally covers the following elements:

- Description of the data sets and their organization;
- Methods and processes for data processing throughout the project;
- Methods of dissemination and sharing;
- Archiving and preservation strategies.

DMP OPIDoR is a planning aid tool for implementing research data management. It enables the production of machine-readable and executable data management plans. Developed by Inist-CNRS, it is made available to the entire French scientific community in higher education and research, as well as to its partners. This tool is one of the components of Recherche Data Gouv, the national ecosystem dedicated to the management, sharing and openness of research data.

To promote the adoption of FAIR principles and to respond to the recommendations of the Research Data Alliance, DMP OPIDoR proposes a common structured data management plan model. Recommended by the French funding agencies' open science network, this model is increasingly widely adopted by organizations

of research. It is extensible and customizable, which allows integrating disciplinary specifics as well as the management of codes and software associated with research projects.

C) Joining a sustainable, European and international dynamic

Open science is today a global movement in full structuring, which profoundly transforms the modes of production, dissemination and evaluation of scientific knowledge.

In this landscape, many European and international infrastructures play a structuring role. One of the major initiatives is the European Open Science Cloud (EOSC), which aims to build an interoperable European environment for sharing and reusing research data. The EOSC is not limited to a technical platform: it is a federated ecosystem designed to connect existing infrastructures and facilitate access to scientific resources produced in Europe.

Other initiatives complement this system. GO FAIR promotes the adoption of FAIR principles at international scale, encouraging the implementation of harmonized practices for data management and publication. The Research Data Alliance (RDA) constitutes a global community of researchers, experts and institutions working on the development of recommendations and best practices for data sharing. It plays a key role in defining open standards, fostering interoperability between disciplines and between countries.

In the field of scientific publishing, OpenAIRE helps structure an open ecosystem in Europe by facilitating the discovery, access and linking between publications, data and research projects. Similarly, the Directory of Open Access Journals (DOAJ) plays a central role in the visibility of open access scientific journals, offering a reference directory guaranteeing quality criteria and editorial transparency. In the area of alternative publishing infrastructures, OPERAS supports the development of the humanities and social sciences by promoting open and sustainable services tailored to their specific needs.

Finally, initiatives such as SCOSS (Global Sustainability Coalition for Open Science Services) contribute to the sustainability of open science infrastructures. They financially support essential services to the ecosystem, often led by academic communities, to ensure their stability and long-term development.

In this international context, France fully participates in this cooperation dynamic. It actively participates in numerous initiatives and supports the development of open infrastructures at the European and global level. Among these commitments is the creation of a Franco-Dutch foundation dedicated to the Directory of Open Access Books (DOAB), an international platform for indexing open access scientific books.

All of these measures illustrate the gradual construction of an international open science ecosystem, based on cooperation, interoperability and the

pooling of resources. Beyond technical infrastructures, this is a genuine scientific and political project aimed at making knowledge more accessible, fairer, and more sustainable on a global scale.

To go further:

Open science training courses are offered in the training catalog of the College of Doctoral Schools via ADUM (with hours validated):

2 online sessions in French:

- Discover Open Science (3h)
- Showcasing my thesis and scientific publications with open access (3h)
- Research Data Management: fundamentals (3h)

1 online session in English:

- Open for you! An introduction to open science (3h)
- Promote my thesis and my scientific publications with open access (3h)
- Research Data Management: Fundamentals (3h)

Sources:

<https://www.u-bordeaux.fr/recherche/ambition-scientifique/science-ouverte>

<https://www.ouvrirlascience.fr/plan-national-pour-la-science-ouverte>

<https://bibliotheques.u-bordeaux.fr/appui-aux-chercheurs>

<https://intranet.u-bordeaux.fr/personnels/appuis-aux-missions/appui-la-recherche/gerer-ses-donnees-de-recherche>