EDSP2 Open Science Guide

Principles and definition of Open Science

Contents

A) Principles and definition of Open Science 2
- 1- Why open up science? 2
- 2- The benefits of open science for a PhD student 2
- 3- How can science be opened? 3

B) Expanding open access to publications 3
- 1- Deposit your results in an open archive 3
- 2- How to find out about the journal's policy 5
- 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 institutional HAL portal of the University of Bordeaux 8

C) Structuring and opening up research data 8
- 1- Prepare and document data 9
- 2- How to choose a data warehouse? 9
- 3- Data warehouse tool 10
- 4- ABDo: the Bordeaux Research Data Workshop 10
- 5- The Research Data Gov institutional space of the University of Bordeaux 10
- 6- Link data to publications 11
- 7- FAIR and DMP 11

D) Be part of a sustainable, European and international dynamic 13

To go further: the training offer in ADUM and the sources 14

A) 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 available free of charge online, without subscription;
- Sharing of research data: the data produced by researchers can be reused and verified;
- Transparency of methods and codes: the protocols, software and tools used are open and reproducible;
- Collaboration and wider 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.

1- Why open up science?

Three main factors should encourage a wider practice of open science. They are scientific, economic and citizen-oriented.

- Scientific: a science that shares the products of scientific research allows for the circulation of protocols, data and results. Interaction between researchers, reproducibility (and therefore transparency and scientific integrity) and reinterpretation are promoted. In addition, openness can also, thanks to (among other things) new data processing tools, open up new avenues of research and promote innovation.
- Economic: A more open science aims to offer an alternative model to the current economic model of scientific publishing. Indeed, the constant rise in costs is no longer sustainable even though digital tools now allow scientists to regain control of the publishing circuit.
- Citizen: A more open science meets citizens' expectations for transparency, and makes it possible to fight against the false scientific information that is currently circulating in the highest spheres of decision-making bodies.

2- The benefits of open science for a PhD student

- Increased visibility and impact
  Open access publications and data are more cited and disseminated. For a doctoral student who is building his or her scientific reputation, it is an opportunity to make his or her work known well beyond his or her discipline or institution, from the beginning of his or her career.

- Enhanced credibility and scientific integrity
  Sharing your data, protocols and source codes demonstrates the rigor of her approach. It also protects the doctoral student in the event of a dispute: everything is traceable and verifiable, which reinforces the community's confidence in his or her results.
- Acceleration of thesis progression
  Free access to the data and publications of other researchers avoids "reinventing the wheel". The doctoral student can rely on existing resources, reproduce analyses and save precious time in his own research.
- Collaboration and networking opportunities
  Making one's work accessible attracts the attention of other researchers, potentially in complementary fields. This can lead to unexpected collaborations, co-publications or invitations to conferences — all assets for an academic or industrial career.
- Preparation for institutional and funding requirements
  More and more funding agencies (ANR, Horizon Europe, etc.) and universities are imposing open access deposit as a condition for obtaining funding. Training in these practices from the doctoral level means anticipating obligations that will be essential throughout one's career.

3- How can science be opened?

3 major axes:

Expanding open access to publications
Structure and open research data
Be part of a sustainable, European and international dynamic

B) Expanding open access to publications

1- Deposit your results in an open archive

There are three main ways to meet this objective: open archive repositories (*the green way*), open access publication (*the golden way*), and publication in open access journals (the diamond way).

The greenway

The green route consists of researchers depositing a version of their scientific publications in an open archive, independently of their publication 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: Established in 1991 by Cornell University, arXiv is considered one of the oldest open archives. It covers the fields of research 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 full-text access to articles in the biomedical and life sciences fields.

The Golden Way

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

- Full open access (or "free") journals: These are peer-reviewed journals in which all articles are accessible free of charge as soon as they are published. Readers do not pay to access the content, and authors generally do not pay publication fees (e.g. OpenEdition, epi-journals or Open U Journals). The Directory of Open Access Journals (DOAJ) lists these journals according to criteria of quality, editorial transparency and respect for copyright.
- The author-pay model (APC): In this model, open access publishing is funded by fees called Article Processing Charges (APCs), paid by the authors, their laboratory or institution. The article is then immediately accessible free of charge to all readers.
- The hybrid model: These are journals mainly funded by subscriptions, but which offer authors the opportunity to make their article open access in exchange for the payment of APC. This means that in the same journal some articles are accessible only by subscription and others are open access. This model is often criticized because it can lead to double payment: institutions pay both subscriptions to read journals and APCs 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 the ANR or Horizon Europe.

The Diamond Way

The diamond route differs from the green and gold routes in its economic model. Unlike the green route, the diamond route directly concerns publication in open access journals. It also differs from the classic golden route because it is not based on the payment of publication fees (APC) by authors or their institutions.

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

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

2- How to find out about the journal's policy

There are several tools that can be used to check a publisher's open access policy before publishing a research paper.

- SHERPA/RoMEO : this service developed by the University of Nottingham lists publishers' policies on open access and classifies them according to different levels of distribution and archiving authorizations.
- WillO : a service developed by the University of Lille, which allows you to know the rights between an author and his or her publisher in terms of open access for a given article.
- Directory of Open Access Journals (DOAJ): the DOAJ lists open access scientific journals according to several criteria: free access to the full text, transparency of the editorial policy, long-term archiving, use of permanent identifiers, respect for copyright and use of Creative Commons licenses.

The DOAJ also provides access to the metadata of articles, improving the visibility of scientific content. Its advanced search mode allows you to find fully open access journals, with APC or hybrid.

3- Predatory journals

Predatory journals are scientific journals that prioritize commercial objectives to the detriment of academic standards and scientific ethics.

Although predatory publishers predate the advent of open access, their recent proliferation has been fostered by the emergence and success of the author-pays model. These publishers exploit the model of publishing fees charged to authors in order to generate profits.

These journals publish articles without a rigorous peer review process, despite the guarantees posted on their websites. They generally offer very fast publication times, low publication costs, a non-existent or superficial review process, as well as editorial boards that are not very transparent or composed of researchers who are difficult to identify. Their article acceptance rate is also very high.

Predatory journals frequently solicit researchers by sending massive unsolicited emails (spam) and particularly target young researchers, especially after the publication of their thesis.

4- Open access rights

The Digital Republic Act (2016)
Article 30 of Law No. 2016-1321 of 7 October 2016 for a Digital Republic (known as the "Lemaire Law"), coded in Article L. 533-4 of the Research Code, 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 published at least once a year, as long as the research has been 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 website;
- This right authorises the deposit of the version accepted for publication (postprint/AAM), with a maximum embargo of 6 months (STM) or 12 months (SHS), without the publisher being able to oppose it.

Plan S and the acceleration of open access

Developed 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 open access scientific publications. In particular, it states that: From 2021, all scientific publications resulting from research funded by public or private grants awarded by national, regional and international funding bodies must be published in open access journals or platforms, or be immediately available without embargo in open repositories.

This plan includes 10 principles available on https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/

The obligations of the funders: ANR and Horizon Europe

In addition to the right opened up by the 2016 digital law, the main research funders have made open access deposit mandatory as part of their funding conditions.

- Agence Nationale de la Recherche (ANR): since the 2022 calls for projects (CFP), all publications from ANR projects must be available in immediate open access under the CC-BY license. Authors must deposit the full text (publisher version or manuscript accepted) in the national open archive HAL indicating the decision code of the ANR project (e.g. ANR-22-CE64-0001), in order to allow automatic monitoring of publications. The Non-Assignment of Rights Strategy (NCSS) must be applied at the time of submission. In addition, a Data Management Plan (DMP) is required within 6 months of the start of the project, and a final version is expected at the end of the project.
- Horizon Europe **:** The European Framework Programme for Funding mandates immediate open access (without embargo) to all peer-reviewed scientific publications, under an open license (CC-BY for articles). A copy of the accepted publisher version or manuscript must be deposited in a trusted repository upon publication. The management of research data according to the FAIR principles is mandatory, and a Data Management Plan (DMP) must be provided within the first 6 months of the project, and then regularly updated. Open access costs (APC for fully open journals) are eligible for funding provided they have been budgeted in the proposal.

Which version should be filed?

- 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 refusal of publication. This version can often be freely deposited in an open archive. The rules of each publisher can be checked on SHERPA/RoMEO.
- Author Accepted Manuscript (AAM): This is the final version of the manuscript accepted by the publisher after peer review, but without the formatting of the journal. The Digital Republic Act authorises its deposit in open access, subject to compliance with any disciplinary embargoes.
- Publisher version: This is the final version published with the publisher's layout. The law does not allow it to be deposited in an open archive, unless the publisher has explicitly authorized it in its policy.

Creative Commons licenses

Scientific works deposited in open archives are 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 in order to specify the conditions of dissemination and reuse.

Creative Commons licenses are free licenses that define the uses authorized by the author for any intellectual work. The author can authorize the copying, distribution and, depending on the case, modification of the work. Creative Commons 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 repository like HAL, before and after peer review, regardless of the publication review.

Recommended by the French National Research Agency since the 2022 calls for projects, it consists of informing the publisher, as soon as the article is submitted, that a CC BY open license is applied to the different versions of the manuscript up to the accepted version. It is also possible to associate it with the Etalab 2.0 license, recommended by the French State and legally equivalent to the CC BY. This system allows immediate open access deposit in HAL as soon as it is published, without embargo. This strategy is supported by France Universités and the CNRS.

5- ORCID Identification

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

The ORCID identifier avoids confusion related to homonyms between researchers. It also includes the different forms of identity under which an author can publish (variations of the name, initials, name changes, etc.).

It links together the scientific work produced by a researcher, even when it is distributed in different databases or information systems. It also facilitates interoperability between platforms by avoiding the repetitive re-entry of personal information and bibliographic references.

The ORCID can also be linked to the IDHAL, the author identifier used in HAL. This connection facilitates communication between the two systems, allows the synchronization of information and prevents researchers from re-entering the same data several 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 at the Bordeaux site, as well as other scientific teams, the opportunity to create and publish journals by ensuring the entire editorial process. This includes the submission of articles, their peer review and their publication online in open access.

A scientific committee examines each request for the creation of a journal in order to verify its integration into the system. Journals must comply with the eligibility criteria recommended by the Committee for Open Science.

7- The institutional HAL portal of the University of Bordeaux

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 accessible their publications in open access. The university's libraries offer personalized support for deposits (verification of rights, quality control of records, HAL workshops) via contact: info@oskar-bordeaux.fr for publications.

C) Structuring and opening up research data

Making research data accessible and reusable as widely as possible responds to several major challenges.

- Fostering innovation and collaboration: enabling the reuse of data for potential new applications and encouraging collaborations between data producers and users.
- Ensure data preservation and reduce costs: Preserve data that is unique or expensive to produce, avoid the costs of duplicating existing data, and reduce the risk of data loss.

- Ensuring scientific transparency: facilitating the replication and replication of research, enabling verification and accuracy of results, and helping to discourage scientific fraud.
- Valuing the work of researchers: generating direct recognition for the researchers who produced the data, in particular through the publication of *data papers*[1] or by adapting citation practices. Some journals not only encourage the citation of datasets, but also integrate them into the main bibliography of scientific articles.
- Meeting institutional and regulatory obligations: applying the requirements of scientific journals or funding agencies, such as those of the European Commission.

1- Prepare and document data

The preparation and documentation of research data, including through metadata, is essential to ensure that it is understood and reused. The Data Management Plan (DMP) helps to organize this work throughout the project.

Data sharing must take into account the legal constraints and policies of the institutions. It is recommended to avoid *the supplementary data* offered by publishers, which are often not well suited to good documentation and effective reuse.

It is preferable to deposit the data in institutional, disciplinary or generalist repositories, which guarantee their durability, accessibility and reuse. Finally, the link between data and publications must be anticipated from the beginning of the project.

2- How to choose a data warehouse?

The recommended approach to sharing research data is to:

- Favor a disciplinary data warehouse when one exists, e.g. PANGAEA, 4TU. ResearchData, NAKALA or Harvard Dataverse;
- Failing that, use a generalist warehouse such as Zenodo or B2SHARE;

The choice of a data warehouse is based on several criteria, including the supported formats, the indexing possibilities and the terms of access to the data (free, restricted or embargoed access). The use of certified warehouses is also an important criterion in order to guarantee the quality, durability and reliability of the service offered. The re3data.org portal facilitates the identification of these warehouses thanks to its "Certificates" filter.

[1] A *data paper* is a scientific publication whose main objective is to describe a research dataset. Unlike a classic scientific article, it offers a detailed description of a dataset and the associated metadata, as well as the conditions and methods of collection, without presenting scientific results based on their analysis.

3- Data warehouse tool

The documentation department of the University of Bordeaux provides a decision-making tool freely accessible online entitled "Finding a data warehouse". This tool makes it possible to identify the generalist data warehouse best suited to the needs of researchers based on four main criteria:

- The need to restrict access to data;
- The existence of free deposit options;
- The possibility of freely defining the conditions of use of the data;
- Proofreading of the deposit before it is put online.

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

4- ABDo: the Bordeaux Research Data Workshop

ABDo (Atelier Bordelais des Données de recherche) is a local service dedicated to researchers, doctoral students, engineers and research support staff from all the institutions on the Bordeaux site. Accredited by the Ministry of Higher Education and Research, ABDo offers a one-stop shop (address: donnees-recherche@u-bordeaux.fr) for: training and awareness-raising on research data management and the FAIR principles; support for the deposit, storage, sharing and archiving of data; legal and ethical advice (intellectual property, licensing, GDPR); assistance in drafting Data Management Plans (DMP); and networking for the exchange of best practices between peers.

5- The Research Data Gov institutional space 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 the sharing of research data (https://entrepot.recherche.data.gouv.fr/dataverse/univ-bordeaux). Based on the free software Dataverse and developed in particular by INRAE, this repository is a preferred solution for the deposit of institutional data, for units wishing a sovereign national solution. Accessible to all members of the university via their UBx identifier (SSO), it allows datasets to be deposited or grouped by structure, team or project. (contact: donnees-recherche@u-bordeaux.fr).

Regardless of the trust warehouse chosen for data sharing, it must ensure a minimum set of functionalities ensuring the dissemination, traceability, storage and re-use of research data. These functionalities also contribute to compliance with the FAIR principles[2].

[2] The FAIR principles are based on four dimensions: Findable , Accessible , Interoperable , and Reusable .

A data warehouse must make it possible to:

- The assignment of a *Persistent Identifier* (PID) of the DOI type, making it possible to ensure the citation, traceability and interconnection of data with other scientific productions, 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 provide a legal framework for the reuse of data, in compliance with French and European law;
- A minimum data retention period consistent with the retention policy defined by the home institution.

6- Link data to publications

Several solutions can be used to establish a link between a scientific article and the data associated with it before the publication of the article. The use of permanent identifiers thus facilitates the sustainable interconnection between publications and research data. Similarly, the referencing of publications associated with data (including *data papers*) is generally possible in most data warehouses, even after the initial deposit of datasets.

After an article has been published, adding a link to the associated data may be difficult, if not impossible, on some editorial platforms. An alternative is to fill in this data in the version of the article deposited in an open archive such as HAL, which allows permanent identifiers to be added in the "Associated data" field. This makes it easier to link publications to datasets.

7- FAIR and DMP

The FAIR Principles

The FAIR Principles aim to promote the discovery, access, interoperability and reuse of research data. They define a set of characteristics that data and their metadata must have in order to facilitate their exploitation by humans and 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 FAIRisation of data and software. The Top *10 FAIR Data & Software Things* resource presents ten key actions for different disciplinary areas to make data and code easier

to find, accessible, interoperable and reusable. Each recommendation is accompanied by practical examples.

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

How to assess the FAIR level of the data?

There are several tools that can be used to assess the extent to which data comply with the FAIR principles:

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

The latter offers a series of questions for each principle, with answers in the form of drop-down menus, allowing for a structured and reproducible evaluation.

The DMP

The DMP (Data Management Plan), or DMP (Data Management Plan), is a summary document that makes it possible to organize and anticipate the entire life cycle of research datasets. It describes, for each dataset, the management methods 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 the stages of the data life cycle" and that specifies "how the data of a project will be managed, from its creation or collection to its sharing and archiving" (INIST-CNRS, 2018, Data Management Plan: summary sheet, DoRANum).

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

- The description of the datasets and their organization;
- The methods and processes of data processing throughout the project;
- The methods of dissemination and sharing;
- Archiving and retention strategies.

DMP OPIDoR is a tool to help with the planning and implementation of research data management. It makes it possible to produce structured data management plans that can be used by machines. 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 opening of research data.

In order to promote the adoption of the 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 open science network of French funding agencies, this model is increasingly adopted by research organizations. It is extensible and customizable, which makes it possible to integrate

disciplinary specificities as well as the management of codes and software associated with research projects.

D) Be part of a sustainable, European and international dynamic

Open science is now a global movement in full structuring, which is profoundly transforming the ways in which scientific knowledge is produced, disseminated and evaluated.

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 the sharing and reuse of research data. The EOSC is not just 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 the FAIR principles internationally, encouraging the implementation of harmonized practices for data management and publication. The Research Data Alliance (RDA) is a global community of researchers, experts and institutions working to develop recommendations and best practices for data sharing. It plays a key role in defining open standards, fostering interoperability across disciplines and countries.
In the field of scientific publishing, OpenAIRE contributes to structuring an open ecosystem in Europe by facilitating 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, by offering a reference directory guaranteeing quality and editorial transparency criteria. In the field of alternative publishing infrastructures, OPERAS supports the development of the humanities and social sciences by promoting open and sustainable services adapted 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 ecosystem services, often supported by academic communities, in order to ensure their stability and long-term development.

In this international context, France is fully involved in this dynamic of cooperation. It actively participates in numerous initiatives and supports the development of open infrastructures on a European and global scale. Among these commitments is the creation of a Franco-Dutch foundation dedicated to the Directory of Open Access Books (DOAB), an international platform for referencing scientific works in open access.
All of these mechanisms illustrate the progressive construction of an international ecosystem of open science, based on cooperation, interoperability and the pooling of resources. Beyond technical infrastructures, this is a real scientific and political project aimed at making knowledge more accessible, more equitable and more sustainable on a global scale.

To go further:

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

2 sessions in French at a distance:
- Discover open science (3h)
- Promote my thesis and scientific publications with open access (3h)
- Research Data Management: Fundamentals (3h)

1 session in English via distance learning:
- 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