

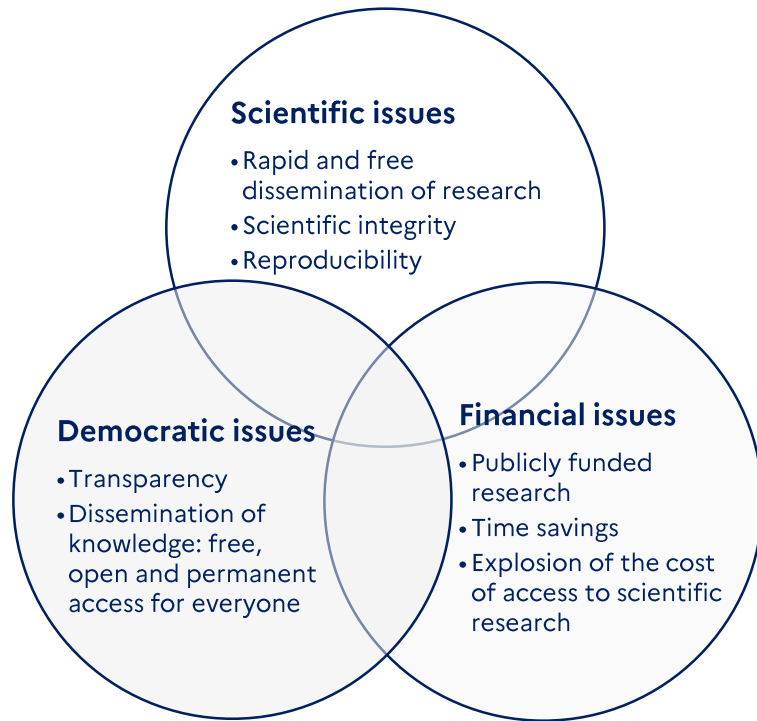
Open science in your PhD : solutions and practical guidance

Alicia León y Barella
Mónica Michel Rodríguez



Open science : context

What's open science?



Free dissemination of results, methods and products from scientific research.

(Site Ouvrir la science – MESR:
<https://www.ouvirlascience.fr/initiez-vous-a-la-science-ouverte/>)

⇒ Sharing among researchers and among disciplines, but also with society as a whole

⇒ Research funded by public money must make its results available to the public

Support with publications

Open Access to publications : Three main “routes”

Disseminating in an open repository: Green Open Access

Self-archiving of scientific works by the researcher or by a person who does it for them

Making the full text available, with or without an embargo period

Dissemination and consultation are done in open repositories (also called open archives)

LillOA
Lille Open Archive

HAL
science ouverte

Commercial open-access publishing: Gold Open Access

The publisher makes the publication available in open access

Author pays.

The author pays for the publication of the article via the APC (Article processing charges)


BREPOLS


MDPI


ELSEVIER

Institutional open-access publishing: Diamond Open Access

Open Access publication funded by a state, university, consortium of public institutions or non-profit organization.

The publication and the dissemination are completely free for both the reader and the author.

Peer Community Journal

 **OpenEdition**
Journals

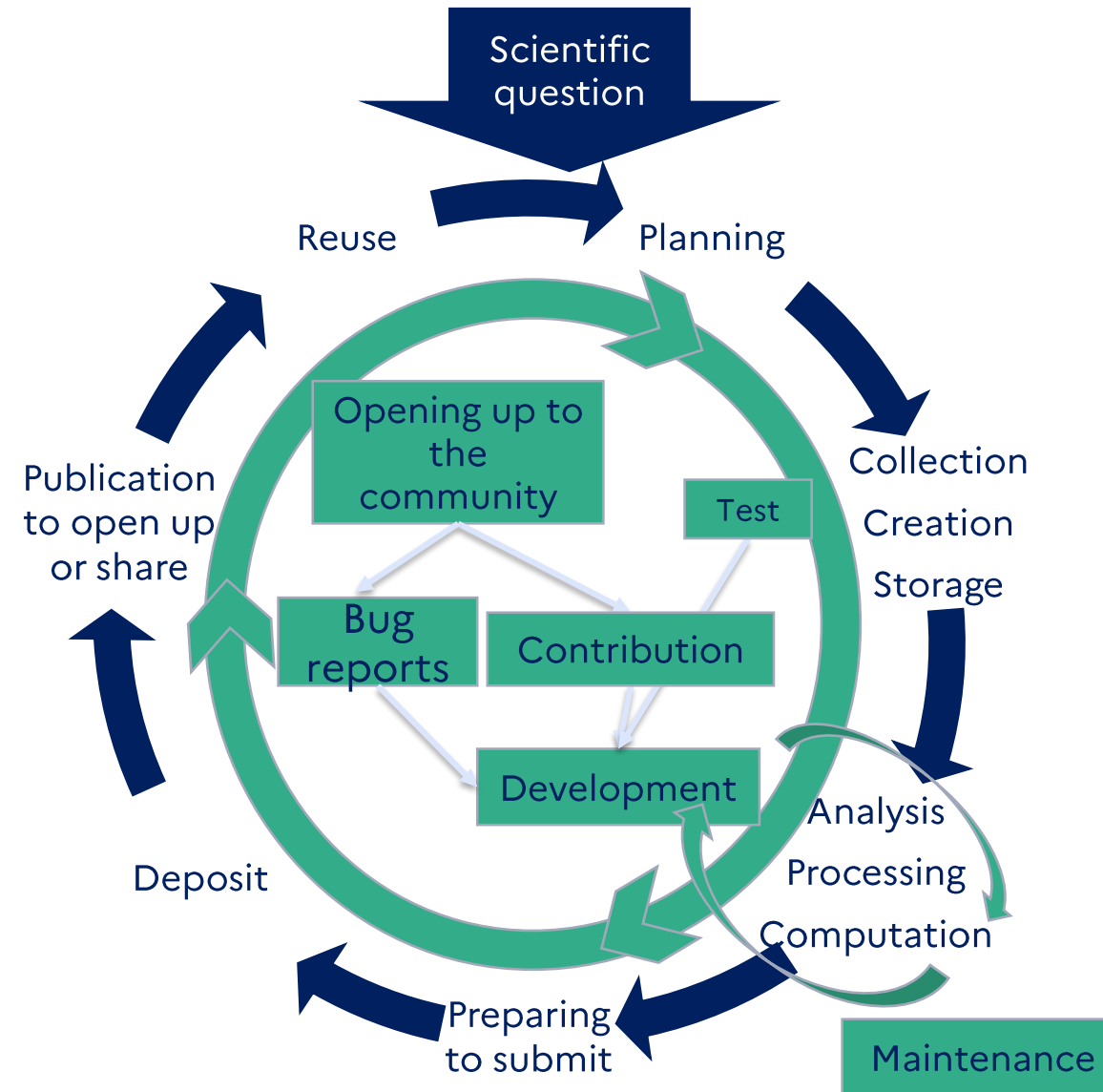

Péren
Plateforme d'édition de revues numériques

What help can I get ?

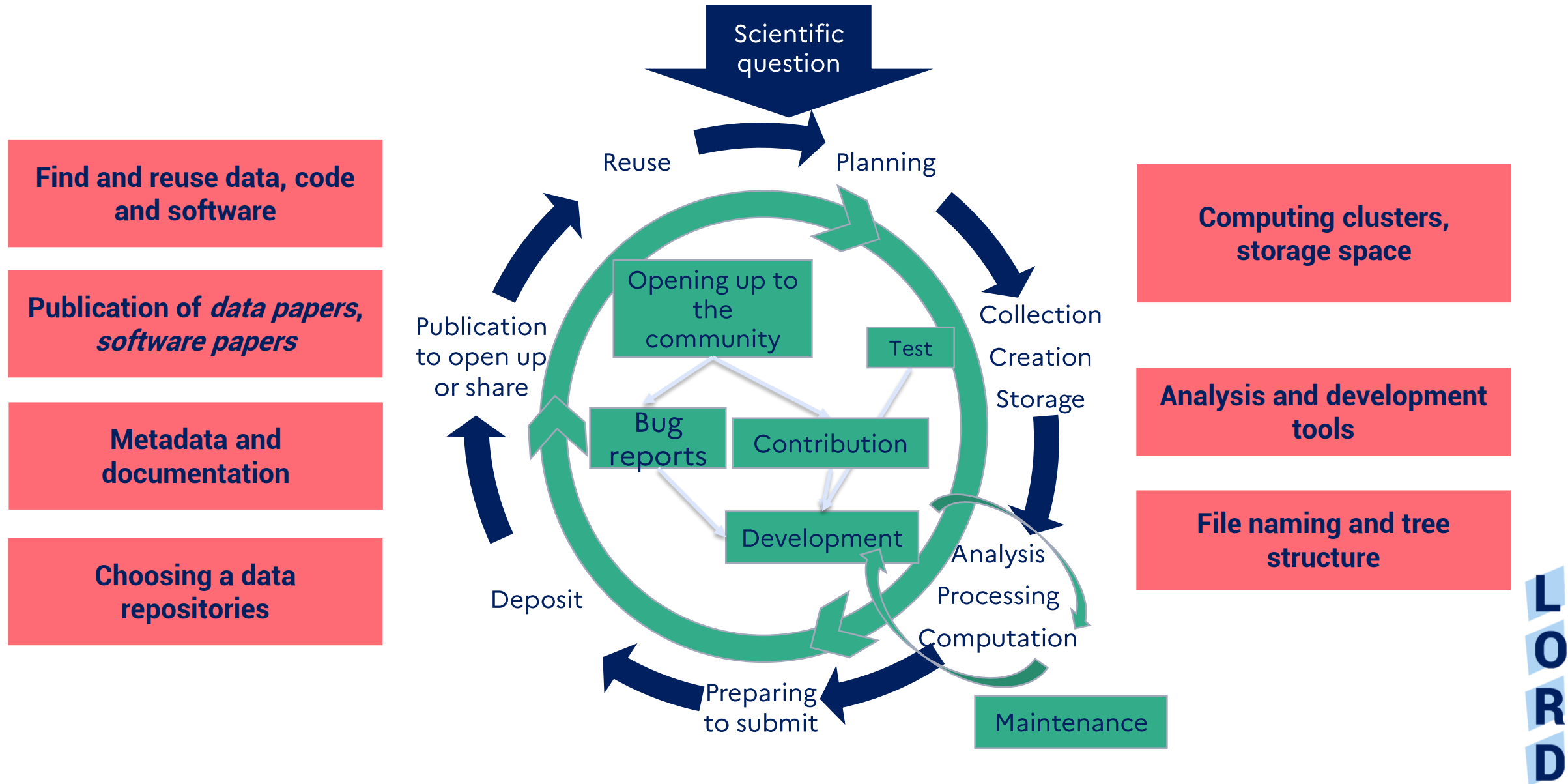
- ◆ Answer questions regarding open science publishing
- ◆ Help you define your publication strategy
- ◆ Give advice on the legal aspects of disseminating your scientific productions
- ◆ Help you pay for Article processing charges ([APC support fund](#))
- ◆ Show how to deposit your publications into an open archive repository
- ◆ Etc.

Support with data, source code and software

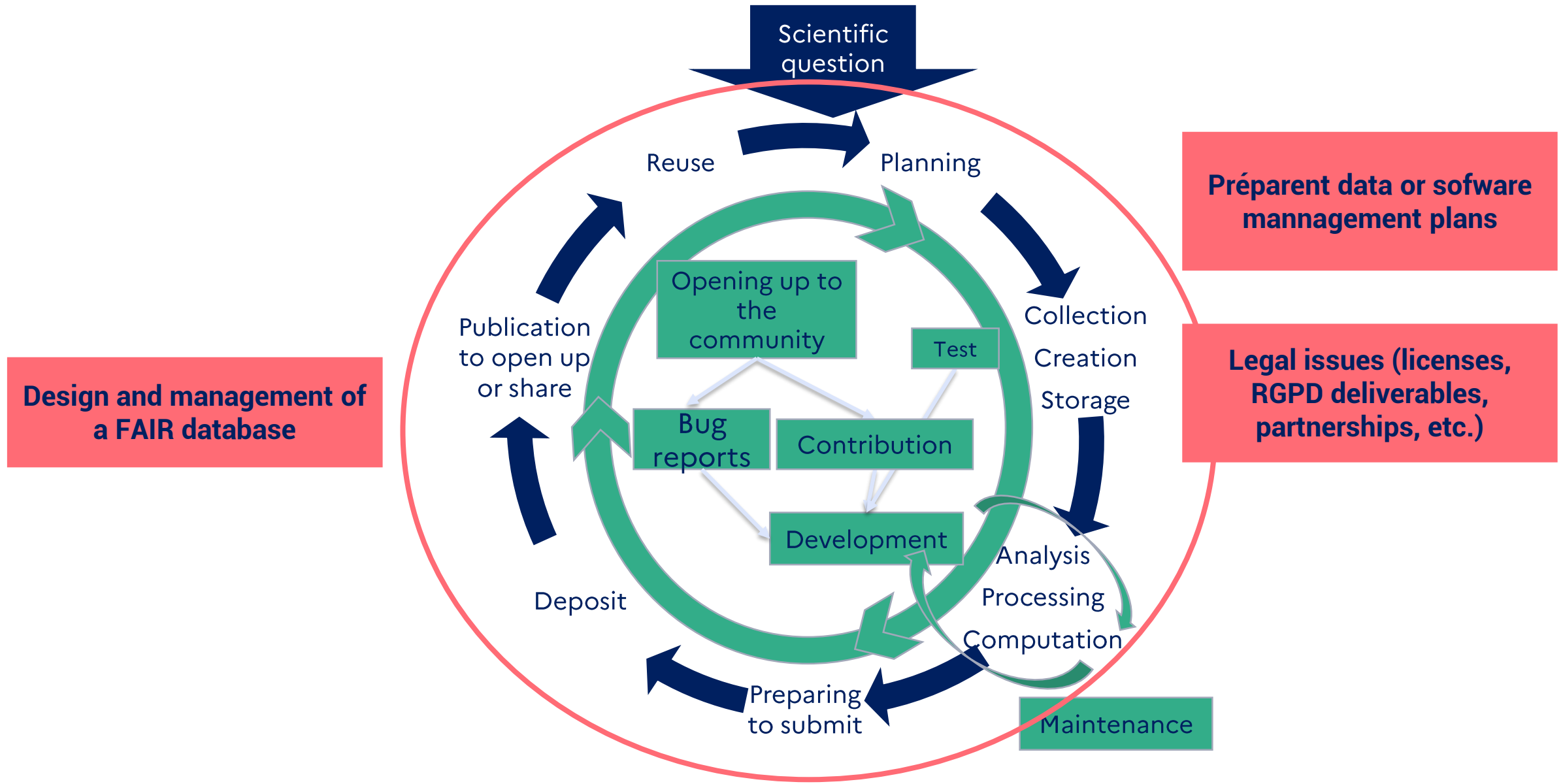
Needs throughout the data, code and software lifecycle



Needs throughout the data, code and software lifecycle



Needs throughout the data, code and software lifecycle



Data management clusters: generalist services throughout France



**Lille
Open
Research
Data**

- ◆ Labellisation en 2022
- ◆ Un projet omni-disciplinaire, inscrit dans son territoire
- ◆ Des services mutualisés, à géométrie variable en fonction de l'implication des établissements dans le projet



LORD, a data management cluster designed for Lille research teams

Lille Open Research Data

- ◆ Within the [Recherche Data Gouv](#) ecosystem
- ◆ Lighten the "data burden" for research teams
- ◆ Pool skills and expertise within and between establishments
- ◆ Build customized support around research data, codes and software
- ◆ Local professional networks close to your research teams

Contact us !

How to get trained

Open science webinars



- ◆ Open data and data management
- ◆ Disseminating my data into Lillodata
- ◆ Open access to publications
- ◆ Program 2024/2025 has yet to be published – stay tuned 😊

<https://bu.univ-lille.fr/chercheurs-doctorants>

Open science doctoral training sessions

Six themes to explore various aspects of Open Science:

- ◆ Open science: general and legal aspects
- ◆ Writing and distributing your thesis
- ◆ Publishing and distributing your work
- ◆ Making your work visible
- ◆ Data, code and software
- ◆ Citizen science



- ◆ 50 sessions of 1 to 3 hours
- ◆ 8 sessions in English
- ◆ ECTS credits awarded
- ◆ Registration on ADUM

<https://bu.univ-lille.fr/chercheurs-doctorants>

Fabrique de la science ouverte



*Ouvrir la recherche :
place à la pratique !*

**la Fabrique
de la science
ouverte**

Workshops to learn how Open Science is made :

- ◆ Events during the year
- ◆ Experts to answer your questions in direct with your scientific production
- ◆ Solutions and platforms for your disciplines.

Next event :

Digital sobriety and research data: let's get practical!
24 June 2024, LILLIAD

<https://fabso.univ-lille.fr/>

Resources to get introduced



MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR
ET DE LA RECHERCHE

*Liberté
Égalité
Fraternité*

 **Ouvrir
la science !**

 **Université
de Lille**

A complete collection with educational content to get introduced you to the issues of open science.



<https://www.ouvrirlascience.fr/passport-for-open-science/>



Passeport pour la science ouverte



Où chercher des ressources ?

Plateformes de revues en accès ouvert : elles proposent des modèles économiques et éditoriaux innovants. Par exemple : **OpenEdition Journals**, le **Revue de géographie de la recherche scientifique** (RGERS) ou **Perse**.

Plateformes de publications ou de documents de travail : elles permettent de partager les travaux les plus récents avant même pendant qu'ils sont évalués par des relecteurs. L'exemple : **arXiv** une archive de publications en mathématiques, physique et astronomie.

Bases de thèses ou de travaux académiques : Rendez-vous sur **Theses.fr**, **Open Access Theses and Dissertation (OATD)**.

Plateaux de recherche spécialisés : comme **Yaddo** ou **Y Google Scholar** vous permettent d'accéder aux publications. **Unpaywall** est une extension du navigateur qui signale les versions en accès libre.

Entrepôts de données : pluridisciplinaires ou spécialisés, ils permettent le stockage sécurisé et la diffusion des données. Exemple : **bioRxiv**, **medRxiv**, **arXiv**, **Open Access Theses and Dissertation (OATD)**.

Logiciels : ces plateformes de développement collaboratif de projets logiciels hébergent et rendent accessibles de nombreux codes sources et logiciels.

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).



Archives ouvertes : les chercheurs et chercheurs peuvent y déposer leurs travaux académiques. Les archives sont institutionnelles ou thématiques. Exemple : **HAL**, archive pluridisciplinaire nationale.

Bases de thèses ou de travaux académiques : Rendez-vous sur **Theses.fr**, **Open Access Theses and Dissertation (OATD)**.

Plateaux de recherche spécialisés : comme **Yaddo** ou **Y Google Scholar** vous permettent d'accéder aux publications. **Unpaywall** est une extension du navigateur qui signale les versions en accès libre.

Entrepôts de données : pluridisciplinaires ou spécialisés, ils permettent le stockage sécurisé et la diffusion des données. Exemple : **bioRxiv**, **medRxiv**, **arXiv**, **Open Access Theses and Dissertation (OATD)**.

Logiciels : ces plateformes de développement collaboratif de projets logiciels hébergent et rendent accessibles de nombreux codes sources et logiciels.

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

Logiciels : ces plateformes de développement collaboratif de projets logiciels hébergent et rendent accessibles de nombreux codes sources et logiciels.

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

L'archive logicielle **Software Heritage** propose un et donne accès à l'ensemble des versions des logiciels dans le code source est librement accessible (mais pas nécessairement réutilisable).

Prévoir la gestion des données

LES DONNÉES DE LA RECHERCHE, C'EST QUOI ?

Les données de la recherche sont définies comme « des enregistrements factuels (chiffres, textes, images, sons, etc.) qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche ». (Source : Organisation de coopération et de développement économiques – OCDE 2007)

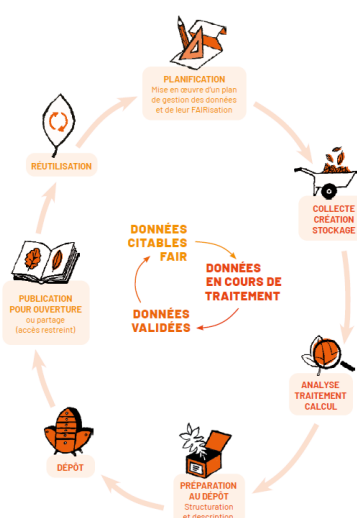
La gestion des données de la recherche concerne l'ensemble du cycle de vie des données, c'est-à-dire des opérations de collecte, production, description, stockage, traitement, analyse, mise en accès et archivage des données.

Pourquoi gérer les données de recherche ?

Une bonne gestion des données est utile pour soi et pour les autres. Elle permet de rendre les données plus facilement trouvables, accessibles, intelligibles par les humains et les machines et réutilisables conformément aux principes FAIR. Voir page 26

Comment gérer les données de recherche ?

La gestion des données s'anticipe dès la conception du projet, grâce à l'élaboration d'un Plan de gestion des données ou DMP (Data Management Plan - DMP en anglais). Ce document vous aide à anticiper puis à mettre en œuvre la manière dont vous organisez les données pendant et à l'issue du projet. Le DMP évolue tout au long du projet : il est souvent édité collectivement (par exemple à l'aide de l'outil **DMP OPIDOR**). Les principales agences de financement de la recherche demandent désormais la fourniture d'un DMP lors de l'appel à projet.



Evaluating research differently

Open science represents a profound change in science and research which means it is currently questioning evaluation practices.

Reinventing peer review

Peer review is a prerequisite for any publication and a guarantee of the reliability of scientific results. The process is usually organised by the journal or the publisher who submits the manuscript to other researchers in the same field as the author and who are not known to the author. However, this system is currently having difficulties particularly because reviewers may be competitors of the author. Conversely, some reviewers sometimes work on research themes which are too far removed from the article needing review. Generally speaking, peer review is not an infallible solution – since the 2000s there has been an increase in the publication of fraudulent or questionable articles because of data manipulation or plagiarism (Grüneisen & Zhang, 2012; Fang, Steen & Casadevall, 2013).

Open science has brought about the emergence of Open Peer Review practices with two main methods: the names of reviewers are made public and/or the review is carried out on a platform that enables all users to comment on the article. This practice has been made easier by the existence of pre-publication platforms like **arXiv** and **bioRxiv** which journals can use to collect comments

Example

The **Peer Community in platform** organises the peer review of pre-publications deposited in an open archive which can lead to the obtention of a certificate of validation. Journals can thus publish articles freely without having to ask for reviewers.

and protocols (registered reports) in registers or journals has led to practices changing. Peer review is carried out in two stages which reduces the effect of publication bias (tendency to publish only positive results) and spotlights the research process.

In the field

Minh-Chia N., PhD student in phonetics, Sorbonne Nouvelle Paris 3 University

I am a phonetician. I study the sounds of languages, particularly those of Vietnamese dialects.

During my master's degree, I made field visits to the province of Phu Tho to collect and record data from the little-known dialect of Kim Thang. I collected, named, phonological and lexical contents, stories, dialogues and songs.

The article has been cited and downloaded but will gradually be replaced by new work. However the primary data is of unlimited value.

With two other researchers, I published an article which is available on HAL. The datasets have been deposited online in the CNRS Pangloss open language sound collection.

I have continued working on my PhD in open science mode. As well as phonetic study, part of my work consists of preparing all the data recorded for long-term conservation in Pangloss. I am happy to contribute in this way to the preservation of my home country's language!

Glossary

APC (Article processing charges)
Charges for publishing immediately in open access which his or her institution). See Open access publications.

Article: preprint or author's version
The version of an article sent to a journal by the authors p

Article: Accepted author manuscript
The version of an article including revisions resulting fr without the publisher's final layout.

Article: version of record
The version of an article with the editor's final layout as i

Data Management Plan (DMP)
An ongoing plan written at the start of a research proje will be managed covering its collection, documentation a sensitive data, conditions for opening or sharing data, et

Data warehouse
These may be multidisciplinary or thematic in one discipl deposited, documented, and disseminated. A warehouse wider access to data than a laboratory server or other lo

Distribution license
The license defines the conditions for distribution and (example: Creative Commons).

Embargo
Period during which a scientific production cannot be disseminated in open access. In the case of state-funded scientific publications, the French Law for a Digital Republic limits the embargo period after which the written work can be openly disseminated regardless of contracts with publishers (see the French law for a Digital Republic). Authors of these may define an embargo period during which the thesis is only available within the academic community.

FAIR (principles)
The aim of the FAIR principles is to make data findable, accessible, interoperable and reusable.

French Law for a Digital Republic
This 2016 law provides a legal framework for depositing certain versions of journal articles in open access repositories if at least half of the funding of the research in question came from the public sector.

By putting research data in the category of public data, this law creates a legal obligation for such data to be freely disseminated.

Act now

When you can, submit your publications to open access journals.

Deposit your publications in an open archive:

- Keep the latest version approved by peers but not yet formatted by the publisher.
- Ask your co-authors for approval.
- Deposit the latest version approved by the peer reviewers in an open archive.

Take part in discussions within your disciplinary community about pre-publications deposited in open archives.

Document and share research data and/or the source code you developed:

- Store data using a perennial system or format in compliance with your team or institution's policy.
- Document the data with metadata so that they are reusable.
- Deposit the datasets associated with your publications in an online repository.
- Deposit your codes in a dedicated perennial open archive like **Software Heritage**.

Follow the evolutions of open science and get involved!

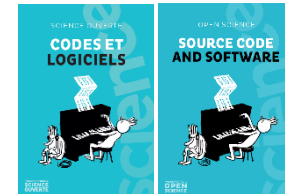


WORTH KNOWING

Would you like to deposit a publication in an open archive or distribute data with a license which allows its reuse? You should talk about this with your co-authors.

DMP OPIDOR facilite l'élaboration de plans de gestion de données et de logiciels.

Livret thématique « Codes et logiciels »



what is software?

The ins and outs of software

Software (or a computer program) is the description, in one or more programming languages, of the way you want to get a computer to process data. **The source code** is the basic element of any software. It is the text written in programming languages by one or more human authors. There are many different languages, each corresponding to specific needs: C, Java, Python, R, Ocaml, Scilab, etc. Computers are only capable of executing elementary instructions, written in a low-level machine language. To execute source code on a computer, you therefore need to:

- either **compile** it, which means converting it beforehand into a functionally equivalent object code written in machine language that can be executed on the computer;
- or **interpret** it, which means translating it on the fly within a dedicated environment (a Python interpreter, etc.). Small source codes intended to be interpreted are often called scripts. Tiny scripts are sometimes called macro-instructions, or macros.

PRATICAL EXAMPLES

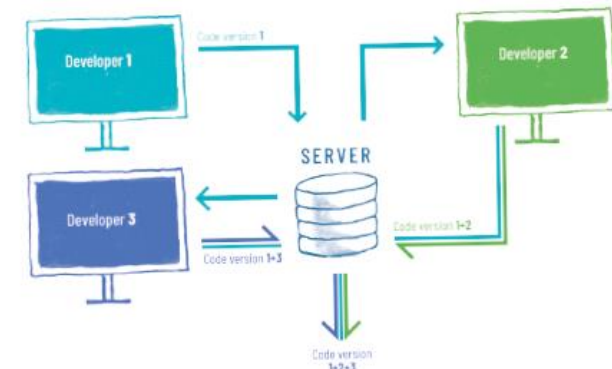
There are different types of licenses, each one adapted to different practices in terms of reuse. **Scikit-learn** is a state-of-the-art statistical learning library that makes the field accessible to all. Designed to be embedded in other software, this library is distributed with a permissive license, the BSD. **WebObs** is a multidisciplinary real-time observation tool used to study natural phenomena. It has been adopted by several countries (Indonesia, Singapore, Peru) to monitor their volcanoes and is distributed under a diffusive license, the GPL 3.0. **Scikit-learn** and **WebObs** are two examples of software that won the Open Science Awards for Research Software (2022).

Organising software's development

Maintaining, organising and developing software of a certain size requires adopting good practices. Evolutions in the source code must be traced using a **version control system**: **Git**, **Mercurial** or **Subversion**.

Many web platforms, known as **software forges** (**bitbucket.org**, **gitlab**, **github**, etc.), simplify these tasks and facilitate the aggregation of communities of contributors. Local instances of these platforms can be put in place by the home institution or by a research infrastructure, such as **HumaNum** for the humanities and social sciences.

Each platform's communities can provide useful benefits: looking for bugs, adding new features, ensuring the software's long-term sustainability, managing contributors and their permissions, implementing quality assurance tests, etc. To benefit from the contributions of a community, one must invest the time and energy needed to organise its functioning: rapidly evaluate and deal with suggestions and contributions made, and plan the roadmap for the software's development.



GOOD TO KNOW

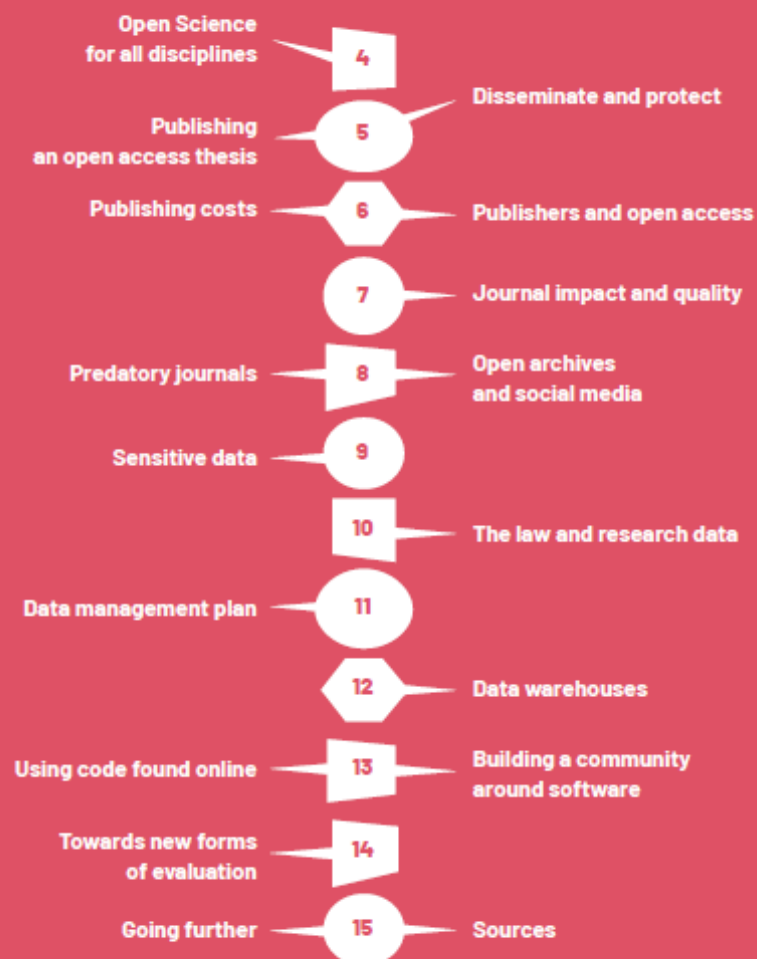
Software forges are not permanent archives: the projects they contain can be modified or deleted. Entire forges can also disappear, as happened to Google Code and Gitorious in 2015, and part of BitBucket in 2020. It is therefore preferable to choose a forge that is managed by a public body and recommended by your institution, and also to ensure that your codes are properly archived over time in dedicated software archives like **Software Heritage**.

priété

ark de
notre
utiliser
verture
se des

Thematic boockclets : *Open Science – Join the debate*

A graphic design inspired on debate and discussion



In my field there is no warehouse to store my research data

There are many data warehouses for specific fields listed in directories like [re3data](#). But not all

disciplines have specialised infrastructure for the publication, open access or sharing of data. If this is true of your field, general or institutional data warehouses are perfectly suited to store your data.

In France, the [Recherche Data Gov ecosystem](#) is a trusted option for any researcher without a dedicated solution.

Whatever the circumstances, the warehouse you choose must be trustworthy, which means it should have criteria guaranteeing reuse of the data: a unique permanent identifier, sufficient description of the data for reuse, use of licences, recognition for the data producers and clear rules governing access, etc.

Find out how your institute can help you.



12

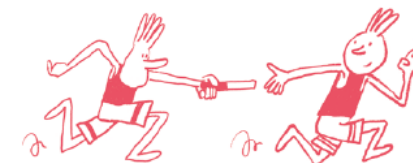
Code available online is royalty-free

The law considers software as intellectual work protected by authors' rights.

And in copyright law, everything that is not explicitly authorised in the software

license is forbidden. In order for a specific action to be permitted, such as copying a software program onto your computer from the Internet, compiling it or executing it, the rights holder must have explicitly authorised it. The conditions of use and reuse are usually set out in a license attached to the software. Be sure to seek it out and adhere to it.

You can consult the guide on Source code and software booklet in the Passport for Open Science collection, which explains how to choose the license best suited to your work.



As freeware specialist François Élie puts it, "you can't pool resources all alone".

In order for people to contribute to your software, there must be an advantage for them: adapting the software to their needs, scientific or technical recognition, etc. If you can deal with contributions as they come in, whether solicited by you or spontaneous, you can also stimulate them by proposing a roadmap and governance approach with which your contributors can identify.

Maintaining a dynamic community requires real investment, something you can delegate in the future. It's a matter of building up long-term trust.

A community will emerge to maintain my software

13

Guide thématique « Données de la recherche »



DONNÉES DE LA RECHERCHE DE QUOI PARLE-T-ON ?

Les données de la recherche dans tous leurs états

Plusieurs définitions existent ; la plus couramment utilisée est celle de l'Organisation de Coopération et de Développement Économique (OCDE) qui définit les données de la recherche comme « des enregistrements factuels (chiffres, textes, images et sons) qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche ».

BON À SAVOIR

Les codes sources et logiciels ne doivent pas être considérés comme des données : ils présentent des enjeux, pratiques et recommandations de partage et d'ouverture particuliers. Consultez le livret **Codes et logiciels** ■

GLOSSAIRE

Catégorie de (méta)données : inventaire des données ou métadonnées destiné à les retrouver.

Curation (dans le cas du dépôt d'un jeu de données) : la curation scientifique consiste à nettoyer, éditer, transformer dans l'objectif d'obtenir des jeux de données « propres », lisibles et plus faciles à traiter. Il existe aussi la curation documentaire et technique qui consiste à vérifier des métadonnées de fichiers de données à déposer dans un entrepôt, dans le but de proposer des modifications et d'améliorer la qualité de description des jeux de données.

Données de la recherche : enregistrements factuels (chiffres, textes, images et sons) qui sont utilisés comme sources principales pour la recherche scientifique et sont reconnus par la communauté scientifique comme nécessaires pour la validation des résultats.

Data paper : publication qui décrit un jeu de données scientifiques, notamment à l'aide d'informations structurées appelées **métadonnées**.

Data Protection Officer (DPO) : personne chargée de la protection des données à l'intérieur d'une organisation.

Données à caractère personnel : données concernant une personne physique qui est identifiée ou identifiable, par exemple par corrélation avec d'autres jeux de données.

Embargo : période pendant laquelle les articles et les données de la recherche déposés sur une plateforme ne sont pas accessibles librement.

Entrepôt de données : service en ligne permettant le dépôt, la description, la recherche et la diffusion des jeux de données. Ils peuvent être pluridisciplinaires ou disciplinaires. Lorsqu'ils respectent une série de critères définis par le guide « Criteria for the Selection of Trustworthy Repositories », ils reçoivent le label de certification qui vise à promouvoir des entrepôts de données fiables et durables.

Identifiant pérenne ou Persistent Identifier (PID) : référence unique et stable pour un objet ou un sujet numérique (un jeu de données, un article, un auteur...). Exemple : Digital Object Identifier (DOI) ou l'identifiant auteur - Open Researcher and Contributor ID (ORCID).

Indexation : attribution à un document de termes distinctifs (des mots-clés par exemple) renseignant sur son contenu et permettant de le retrouver.



Les données de la recherche peuvent être des enregistrements sonores, des images vidéo, des images satellitaires, des images issues de microscopes, un corpus textuel, des transcriptions, un tableau de résultats d'une enquête ou d'un test, des relevés de température d'une série temporelle ou toute autre mesure sur le terrain, le contenu d'une base de données...

Les données de la recherche sont caractérisées notamment par :

- Leur mode d'obtention : données produites dans le cadre d'expérimentation ou d'analyse par des instruments, données d'observation, données...
- Leur format : données avec format ouvert ou propriétaire.
- Leur contexte de production : partenariat industriel, laboratoire en régime restrictif.

POUR ALLER PLUS LOIN

RESSOURCES GÉNÉRALES

Collection Passeport pour la science ouverte.
<https://www.ouvrirscience.fr/passeport-pour-la-science-ouverte/?menu=3>

▼ Ouvrir la science : ressources.

<https://www.ouvrirscience.fr/category/ressources/>

PLATEFORMES D'INTÉRÊT

▼ Doranum propose des ressources pour accompagner la communauté scientifique dans la gestion et le partage des données. Vous y trouvez des contenus d'autoformation par thématiques (enjeux, dépôt, plan de gestion, métadonnées...) ou par disciplines : <https://doranum.fr>

Le Groupe de travail Science ouverte - Données de ▼ Couperin <https://gtso-couperin.org/groupe-donnees/>

▼ Recherche Data Gouv propose des guides, classes virtuelles et tutoriels : <https://recherche-data.gouv.fr/fr/aide-en-ligne>

▼ CoopIST du CIRAD : <https://coop-ist.cirad.fr/gere-des-donnees>

▼ Réseau des URFIST propose des formations sur place ou à distance : <https://agefor.reseau-urfist.fr/#/>

▼ EcoInfo - groupement de services du CNRS sur la réduction des impacts environnementaux et sociétaux négatifs des technologies du numérique : <https://ecoinfo-cnrs.fr/>

▼ Labos'point5 - collectif de membres du monde académique, de toutes disciplines et sur tout le territoire, partageant un objectif commun, celui de « mieux comprendre et réduire l'impact des activités de recherche scientifique sur l'environnement, en particulier sur le climat » : <https://labospoint5.org/>

SUR LE TERRAIN

NAOMI T.
Maîtresse de conférences en linguistique et germaniste à l'Université de Leiden

Presque toutes mes données, articles et réflexions sont aujourd'hui en accès ouvert, mais cela n'a pas toujours été le cas. Il est important de le dire : pratiquer la science ouverte est un processus, et on n'a pas besoin de tout rendre public tout de suite !

J'ai publié ces données en accès ouvert sur ORTOLANG (Outils et Ressources pour un Traitement Optimisé de la LANGUE) des le début de ma thèse et avant même d'avoir publié mes résultats.

De mon côté, cela a commencé par mes corpus annotés. Les données annotées sont issues de débats parlementaires français, allemands et britanniques. Mon projet a ainsi consisté en la mise en valeur de transcriptions de débats parlementaires en France, en Allemagne et au Royaume-Uni en format XML-T1E sous licence CC-BY 4.0 afin de faciliter leur diffusion et réutilisation le plus rapidement et le plus largement possible.

La réutilisation des données parlementaires est un enjeu démocratique de taille : alors que les transcriptions de séances parlementaires sont toutes disponibles sur les sites respectifs des parlements, l'exploitation des données à des fins de recherche reste compliquée.

Cette démarche m'a permis de valoriser plus largement les résultats de ma recherche. Deux data papers décrivent le processus afin de le rendre transparent et reproductible. Je suis très heureuse d'avoir été lauréate du prix Science ouverte des données de la recherche dans la catégorie « réutilisation des données » grâce à ce travail ■

Si je n'avais qu'un seul conseil : lancez-vous !



Données couvertes par la protection du patrimoine scientifique et technique de la Nation, par le secret défense...

Le Code du patrimoine et les dispositions relatives à l'accès aux archives publiques encadrent la durée pendant laquelle ces données doivent rester confidentielles. À l'issue de cette période, les données peuvent être diffusées librement. Cela s'applique par ailleurs aux autres types d'exceptions.



Données couvertes par le secret professionnel...

Les brevets protègent les inventions et non les données sous-jacentes qui ont permis leur avènement. Néanmoins, avant l'obtention d'un brevet, il convient d'être vigilant à ce que la diffusion des données ne conduise pas à une révélation de l'invention. Une fois le titre de propriété intellectuelle obtenu, les données associées à l'invention peuvent être ouvertes, si rien ne s'y oppose par ailleurs.

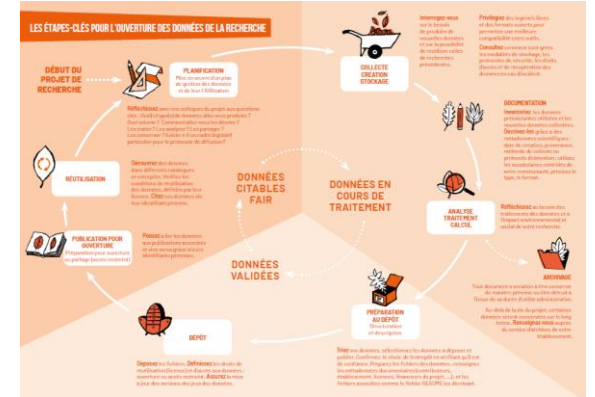
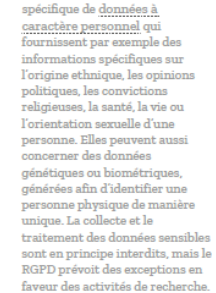
Données produites par des laboratoires situés en zone à régime restrictif...

Ces dernières ne sont pas automatiquement exclues du principe d'ouverture par défaut : pour déterminer les données à garder confidentielles, il convient de se rapprocher des personnes habilitées à se prononcer sur les restrictions de diffusion comme notamment le Fonctionnaire de Sécurité de Défense ou FSD. Ensuite, comme dans tout autre projet, il revient aux équipes d'identifier les données publiques et achevées qui peuvent être ouvertes.

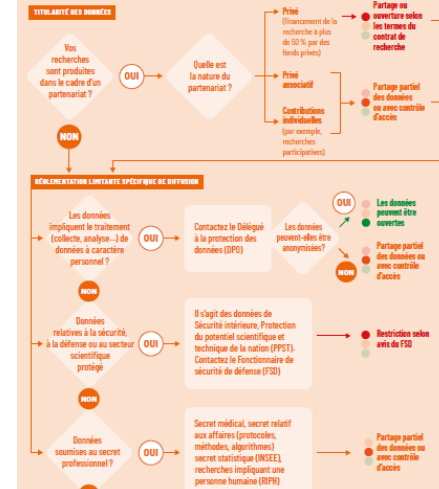
Données à caractère personnel

Leur gestion doit être anticipée au maximum. En effet, pour pouvoir traiter, collecter, enregistrer, modifier, ou même transmettre des données à caractère personnel, il vous faudra contacter le **délégué à la protection des données (DPO ou DPO - Data Protection Officer)** de l'organisme d'appartenance de votre directeur d'unité pour une inscription au registre des traitements (conformément au Règlement général sur la protection des données - RGPD) ou pour une demande d'autorisation à la Commission nationale d'informatique et des libertés (CNIL) ■

Les données sensibles constituent une catégorie spécifique de données à caractère personnel qui fournissent par exemple des informations spécifiques sur l'origine ethnique, les opinions politiques, les convictions religieuses, la santé, la vie ou l'orientation sexuelle d'une personne. Elles peuvent aussi concerner des données génétiques ou biométriques, générées afin d'identifier une personne physique de manière unique. La collecte et le traitement des données sensibles sont en principe interdits, mais le RGPD prévoit des exceptions en faveur des activités de recherche.



Vous avez produit des données dans le cadre de vos activités de recherche, mais vous vous interrogez sur les modalités de partage et diffusion ? Voici quelques questions pour vous guider :



Vous avez répondu NON à toutes ces questions ? Une fois achevées et validées scientifiquement, les données doivent être ouvertes. Le personnel d'appui de proximité peut vous accompagner dans la gestion et la préparation de la diffusion de vos données.



Any questions?

CONTACT

Cheffe du service science ouverte

Alicia León Y Barella

Chargée de projet LORD

Mónica Michel Rodríguez

Alicia.leonybarella@univ-lille.fr

monica.michel-rodriguez@univ-lille.fr

science-ouverte@univ-lille.fr

<https://bu.univ-lille.fr/chercheurs-doctorants>

