



HAL
open science

Un plan de gestion de données pour sa thèse. Pourquoi ? Comment ?

Renaud Delemontez-Sage, Aurélie Gono, Fedor Reshetnikov

► **To cite this version:**

Renaud Delemontez-Sage, Aurélie Gono, Fedor Reshetnikov. Un plan de gestion de données pour sa thèse. Pourquoi ? Comment ?. Doctorat. France. 2026. <hal-05627813>

HAL Id: hal-05627813

<https://univ-paris-dauphine.hal.science/hal-05627813v1>

Submitted on 20 May 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



BU.DAUPHINE.PSL.EU

19 mai 2026

**Un plan de gestion de données
pour sa thèse
Pourquoi ? Comment ?**



Plan

- Données de la recherche et science ouverte
- Principes FAIR et plan de gestion de données (PGD)
- Rédiger un PGD :
 - Au début de la thèse
 - Pendant la thèse
 - À la fin de la thèse
- Valoriser ses données





Tour de table



Données de la recherche et science ouverte

Science ouverte : la définition du Comité pour la Science Ouverte

« La science ouverte est la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique.

Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données, aux codes sources et aux méthodes de la recherche.

Elle permet à la recherche financée sur fonds publics de conserver la maîtrise des résultats qu'elle produit. »

Comité pour la Science Ouverte. [Deuxième Plan national pour la science ouverte](#) : généraliser la science ouverte en France 2021-2024.


MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
*Liberté
Égalité
Fraternité*



Données de la recherche : une définition

Définition de l'OCDE dans : [Recommandation of the Council Concerning Access to Research Data from Public Funding](#) :

« 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 des résultats de recherche . »

Cette définition ne couvre pas : « les carnets de laboratoire, analyses préliminaires ou projets de documents scientifiques, programmes de travaux de recherche futurs, examens par les pairs, communications personnelles avec des collègues, ou objets matériels. »





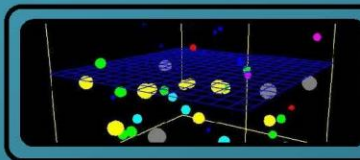
OECD (2020), “Data for science, technology and innovation: Definitions, scope and objectives”, in Enhanced Access to Publicly Funded Data for Science, Technology and Innovation, OECD Publishing, Paris. DOI: [10.1787/3631ee20-en](https://doi.org/10.1787/3631ee20-en)



Données de la recherche : une typologie

Les données de la recherche

QU'ÈS AQUÒ ?

	<p>Données issues de l'observation</p> <ul style="list-style-type: none"> • Issues de l'expérimentation ou dérivées de sources existantes qui sont analysées en vue de produire ou de valider des résultats de recherche originaux • Non reproductibles / Uniques / Irremplaçables 	<p><i>Données d'enquêtes, données de télédétection, données d'échantillonnage, enregistrements sur le terrain, images numériques en microscopie...</i></p>
	<p>Données expérimentales</p> <ul style="list-style-type: none"> • Créées en laboratoire sur des équipements • Reproductibles 	<p><i>Séquences de gènes, chromatogrammes, résultats d'essais agronomiques, de sélection végétale ou animale...</i></p>
	<p>Données dérivées ou compilées</p> <ul style="list-style-type: none"> • Bases de données issues de la compilation d'un ensemble de données collectées et/ou créées • Résultats de fouille de texte ou de fouille de données • Obtenues suite au traitement ou à la combinaison de données brutes • Issues d'un traitement, d'une sélection...de données brutes 	<p><i>Data mining, text mining, fichiers de texte...</i></p>
	<p>Données de références</p> <ul style="list-style-type: none"> • Données ayant en général été publiées • Données triées et agrégées • Collection de données 	<p><i>Textes, images d'archives, base de données d'images ou de vidéos...</i></p>
	<p>Données de simulation</p> <ul style="list-style-type: none"> • Générées par des modèles informatiques • Projections 	<p><i>Modèles climatiques, modèles économiques, modèles de croissance des plantes, modèles de production agricole, modèles multi-agents de gestion des ressources renouvelables, modèles d'accompagnement...</i></p>



Source : <https://amubox.univ-amu.fr/s/LSLXnBdJ9SJqZAF>

Des exemples de typologies de données

- **Produites** (à partir d'une machine, d'instruments de mesures, d'un questionnaire...)
- **Collectées** (données existantes)

- **Brutes**
- **Traitées** (formatées, structurées, raffinées)
- **Dérivées** (à partir d'un traitement spécifique)

Disciplinaires

- **Biologie** (Données « omiques », de structures moléculaires, cellulaires, physiologiques, phénotypiques, cliniques, épidémiologiques)
- **Maths ou informatique** (Codes de calculs, codes source, données produites par des logiciels, modèles climatiques ou économiques)
- **Sciences humaines et sociales**
 - *Données d'observation* (relevés de terrain, questionnaires et résultats d'enquêtes, retranscriptions d'entretiens)
 - *Données dérivées ou compilées* (données issues de fouilles de textes, bases de données compilées)

- **Données d'observation** (collectées à l'instant T, 1) nécessitent un appareil descriptif conséquent (conditions, méthodologie, équipement, etc.). 2) Indissociables d'un contexte donné, elles sont uniques et *impossibles à reproduire*.
- **Données expérimentales** (obtenues à partir d'équipements en laboratoire, suivant une méthodologie bien définie. Potentiellement reproductibles, mais à des coûts parfois prohibitifs)
- **Données computationnelles ou de simulation** (issues de simulations à partir de modèles informatiques. Potentiellement reproductibles si le modèle informatique est correctement documenté)

Pourquoi ouvrir les données de la recherche ?

Pour une recherche scientifique plus cumulative

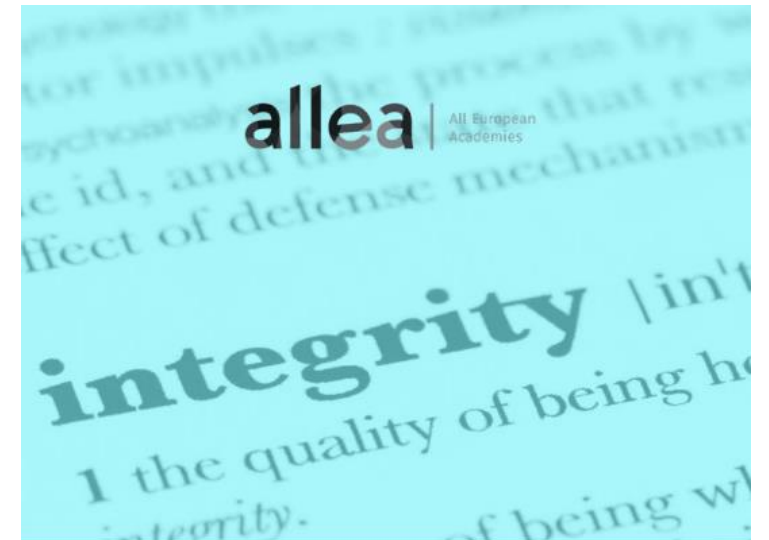
- Gain de temps et de ressources : 80% des données produites ces 20 dernières années seraient perdues
- Partage des résultats « négatifs »
- Réalisation de méta-études

Pour une recherche scientifique plus transparente

- Gage de fiabilité et d'honnêteté
- Lien très fort entre intégrité scientifique et science ouverte, dont les données sont une composante importante
- Enjeu de la reproductibilité / répliquabilité des expériences

Pour une recherche scientifique plus ouverte sur la société

- Permettre la circulation des savoirs et favoriser l'innovation
- Public money ? Open data !
- Science avec et pour la société (SAPS)



**The European
Code of Conduct for
Research Integrity**
REVISED EDITION 2023



ALLEA (2023) [The European Code of Conduct for Research Integrity – Revised Edition 2023 \(French\)](#). Berlin.
DOI: [10.26356/ECOC-French](https://doi.org/10.26356/ECOC-French)



Les principes FAIR et le PGD

Le cycle de vie des données de recherche

Le cycle de vie des données de recherche, c'est l'ensemble des étapes de gestion, conservation, diffusion et réutilisation des données scientifiques liées aux activités de recherche. ([CIRAD](#))

6 étapes :

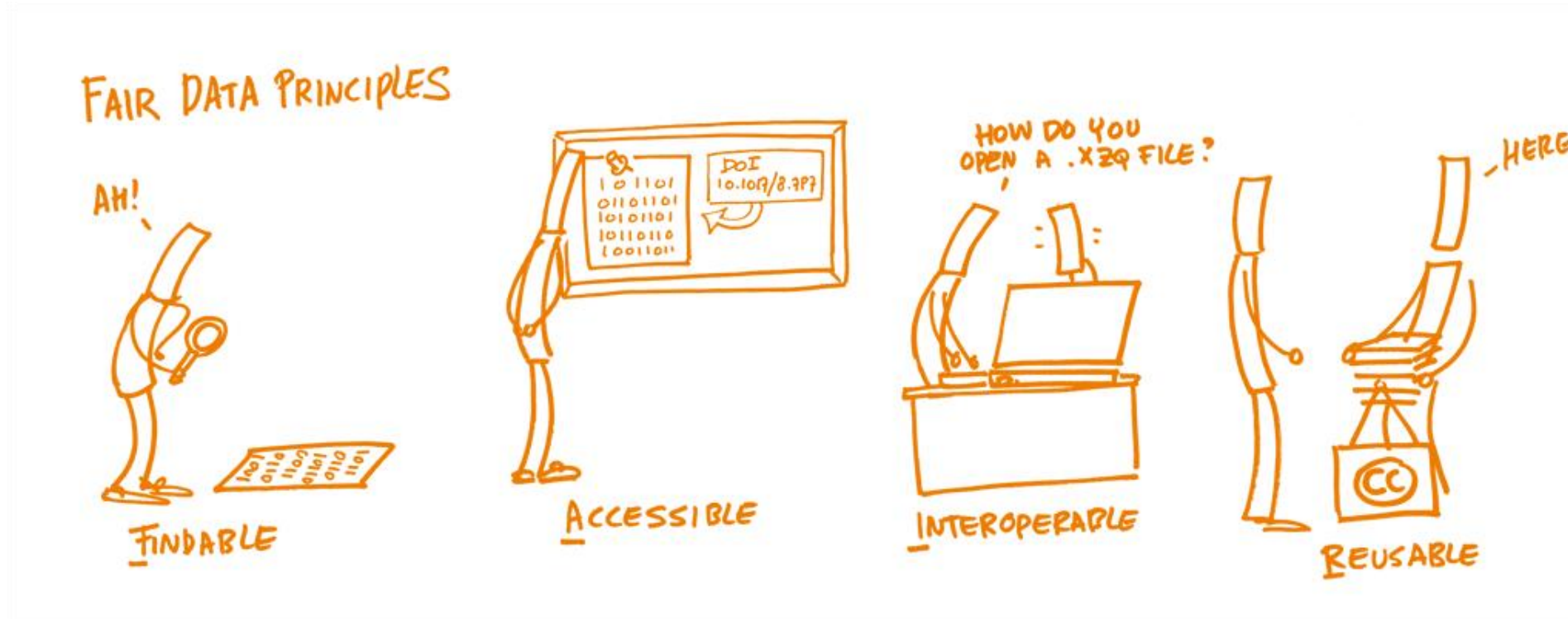
1. Planification
2. Collecte / Création
3. Traitement / Analyse
4. Accès / Partage
5. Préservation
6. Réutilisation



DoRANum (2021). Le cycle de vie des données de la recherche.

DOI : [10.13143/gzj2-j593](https://doi.org/10.13143/gzj2-j593)

Les principes FAIR



F comme
Faciles à trouver

A comme
Accessibles

I comme
Interopérables

R comme
réutilisables

BEZJAK, Sonja, CLYBURNE-SHERIN, April, CONZETT, Philipp, et. al., 2018. [Open Science Training Handbook](#).

Les principes FAIR, comment faire en pratique ?

Le Minimum à FAIR

Cliquer sur les illustrations pour avoir plus d'informations



Publié le : 21/11/2024



Centre de ressources
supports pédagogiques, e-formation
Recherche Data Gouv



Choisir



un **entrepôt de données** adapté dans lequel déposer ses données

Vérifier



que l'entrepôt dans lequel les données sont déposées leur attribue un **identifiant pérenne**

Documenter



un maximum les **données** et utiliser des vocabulaires contrôlés de sa communauté disciplinaire

Définir



clairement les **droits d'accès** aux données au moment du partage

Privilégier



des **formats de fichiers ouverts**, non-propriétaires

Attribuer



une **licence de réutilisation** à ses jeux de données

Mentionner



les **productions liées aux données** (publications, thèses, autres jeux de données, codes et logiciels...)

DoRANum et Réseau des Ateliers de la Données. 2024. « [Le minimum à FAIR](#) ».

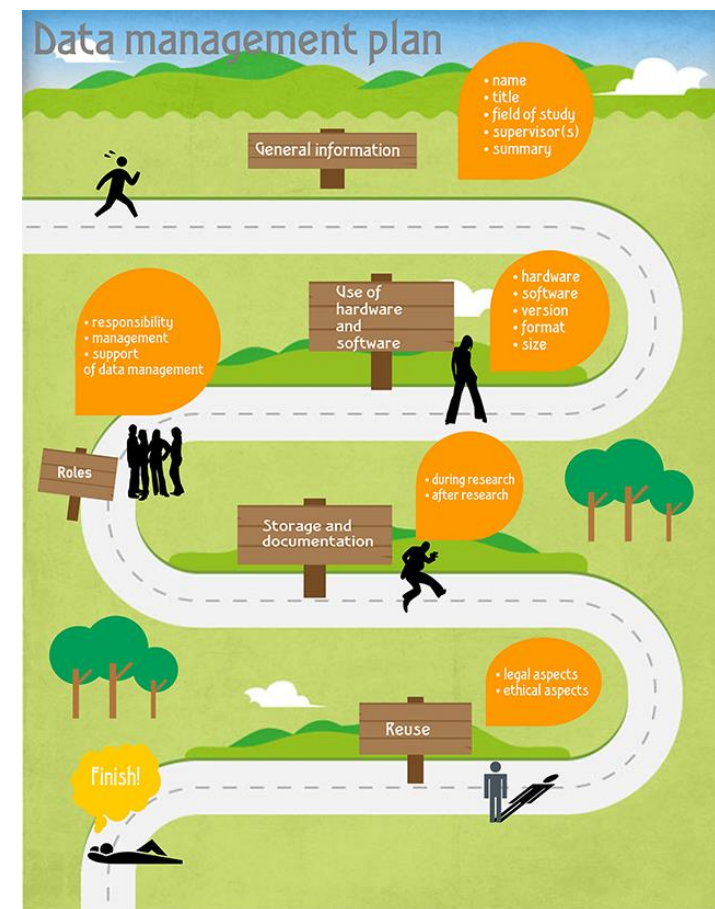
Qu'est-ce qu'un plan de gestion de données ?

Un document formalisé

- Il décrit la façon dont les données produites seront obtenues, documentées, utilisées, gérées et disséminées pendant la phase de recherche et une fois le projet terminé.
- Il est un guide au quotidien tout au long du projet : il aide à retrouver les données après plusieurs années.

Un document évolutif

- À compléter au fur et à mesure du projet.
- À envoyer aux financeurs à différents stades du projet.



SLOUWERHOF, Inge. [LibGuides: Data management for students: Data management plan.](#)

Pourquoi remplir un PGD ?

Pour les chercheurs

- Se poser les bonnes questions avant de se lancer dans le projet
- Mettre en place de bonnes pratiques
- Faciliter le travail à plusieurs ou sur plusieurs sites

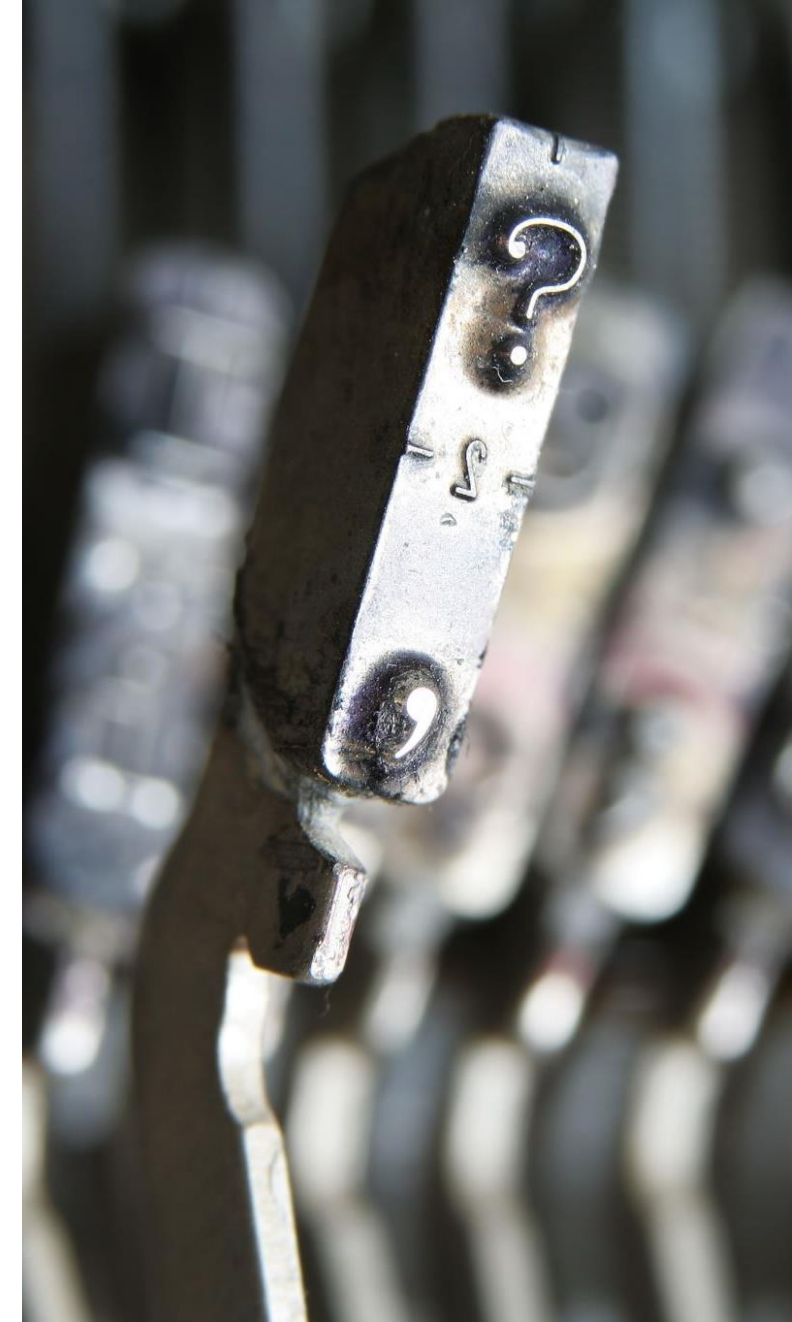
Pour anticiper / planifier

- Éviter des drames (panne technique, faille de sécurité...)
- Évaluer les coûts

Pour répondre aux enjeux de la science ouverte

- Faire en sorte que les données « finales » soient d'une qualité irréprochable, pour tous les usages qu'on veut en faire
- Pouvoir réutiliser ses données à moyen terme

Pour répondre aux exigences des financeurs



Où trouver un modèle de PGD/DMP ?



- Plateforme portée par l'INIST-CNRS, avec un objectif : « faciliter la planification et la mise en œuvre de la gestion des données par le chercheur »
- Possibilité de consulter des [DMP publics](#)
- Fonctionnalités :
 - Un [modèle structuré « Science Europe »](#), lié à des référentiels pour un remplissage facilité ;
 - Des recommandations, qui peuvent être personnalisées par les établissements ;
 - Une gestion des droits de partage et de propriété, permettant de collaborer à plusieurs ;
 - Un outil d'export, pour que le PGD soit lisible par les humains et par les machines.



Le PGD au début de la thèse

Quelles questions se poser au début de la thèse ?

- Quelles données seront produites et comment ?
- Pourrai-je diffuser les données produites dans le cadre de la thèse ?
- Est-ce que les financeurs de la thèse ont des attentes particulières ?
- Comment faire preuve d'éthique et d'intégrité dans le cadre des données de la recherche ?



Réutiliser des données ou en créer de nouvelles ?

Réutiliser des données existantes

- Vérifier les autorisations et licences attribuées
- Pour trouver des entrepôts :
 - [re3data](#) : registre global des entrepôts de données de recherche
 - [Entrepôts de confiance](#) sélectionnés par le CoSo
 - [Entrepôts](#) recensés par DataCite
- Il existe également des bases de données publiques : [Insee](#), [data.gouv.fr](#)...

Créer de nouvelles données

- Renseigner les modèles et la description détaillée des méthodes de collecte, de nettoyage, d'analyse
- Seule la personne à l'origine des données peut effectuer ce travail
- Plus la réutilisation des données est loin du point d'origine, plus il devient difficile de les retrouver et de les interpréter

Quel cadre juridique pour la diffusion des données ?

La loi pour une république numérique, article 30 (2016) - [Article L533-4 - Code de la recherche](#)

« II.- Dès lors que **les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'État**, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne **ne sont pas protégées par un droit spécifique ou une réglementation particulière** et **qu'elles ont été rendues publiques** par le chercheur, l'établissement ou l'organisme de recherche, **leur réutilisation est libre**.

III.- L'éditeur d'un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication. »

[Guide d'application de la Loi pour une République numérique pour les données de la recherche \(2022\)](#)

Par principe, les données produites ou reçues dans le cadre de la recherche publique sont juridiquement considérées comme des « documents administratifs » ou des informations publiques.

Les données de recherche issues de l'activité courante des établissements de recherche et détenues par eux sont ainsi soumises, sauf exceptions, aux **principes d'ouverture par défaut et de libre réutilisation** fixés par le Code des relations entre le public et l'administration (CRPA).

Les exceptions à l'ouverture des données

Exception à l'ouverture de certaines données par l'article L533-4 du Code de la recherche, notamment celles répondant à « un droit spécifique ou une réglementation particulière »

Exemples de données exclues :

- Données portant atteinte au secret de la défense nationale en cas de divulgation ;
- Données privées ou à caractère personnel ;
- Données couvertes par le secret des affaires ;
- Documents réalisés en exécution d'une prestation ;
- Données couvertes par des droits de propriété intellectuelle appartenant à des tiers

ARÈNES, Cécile, MAUREL, Lionel et RENNES, Stéphanie, 2022. [Guide d'application de la Loi pour une République numérique pour les données de la recherche \[en ligne\]](#).

Ministère de l'enseignement supérieur et de la recherche.



Les données personnelles et le règlement général sur la protection des données (RGPD)

- Une « donnée personnelle » est « toute information se rapportant à une personne physique identifiée ou identifiable ».
- Un « traitement de données personnelles » est une opération, ou ensemble d'opérations, portant sur des données personnelles, quel que soit le procédé utilisé.
- Le RGPD s'applique à toute organisation, publique et privée, qui traite des données personnelles pour son compte ou non.



The screenshot shows the CNIL website interface. At the top right, the CNIL logo is displayed with the tagline: "PROTÉGER les données personnelles, ACCOMPAGNER l'innovation, PRÉSERVER les libertés individuelles". Below the logo, navigation links include "MON QUOTIDIEN", "EXERCER MES DROITS", "À TÉLÉCHARGER", and "LA CNIL" with a search icon. The main content area has a breadcrumb trail: "🏠 > Recherche scientifique (hors santé)". The title "Recherche scientifique (hors santé)" is centered. Below the title, a paragraph states: "La recherche scientifique doit s'accompagner de garanties fortes pour protéger les droits des personnes et respecter le RGPD et la loi Informatique et Libertés. Les contenus de cette page ne s'appliquent qu'en dehors du champ de la santé, pour lequel *des règles spécifiques s'appliquent.*"

Données de la recherche et financeurs de la recherche

Au niveau européen

Programme Horizon Europe (2021-2027)

- L'open access aux publications avec comité de lecture est obligatoire.
- L'open access aux données de recherche est obligatoire sous la condition « **aussi ouvertes que possible, aussi fermées que nécessaire** », en suivant les **principes FAIR**.
- Les **plans de gestion de données (PGD)** sont obligatoires pour tous les projets qui génèrent ou réutilisent des données.
- *L'open access* à des productions autres que les publications et données (logiciels, modélisations, applications, etc.) est recommandé.

Au niveau de la France

- Mise en place d'une politique commune pour les agences françaises de financement de la recherche
- Axes de travail :
 - Promouvoir l'accès ouvert immédiat.
 - Soutenir la diversité des modèles de publication en accès ouvert.
 - **Déployer un modèle commun de plan de gestion de données structuré.**
 - Prendre part aux discussions sur la réforme de l'évaluation de la recherche.



Les acteurs de la donnée à Dauphine

Direction du Soutien à la Recherche (DSR)

- Pour tous les aspects contractuels
- dsr@dauphine.psl.eu

Délégué à la Protection des Données (DPD / DPO)

- À consulter dès que l'on manipule des données personnelles ([page ENT](#))
- dpd@dauphine.psl.eu

Comité d'Éthique de la Recherche

- Délivre un IRB (Institutional Review Board)
- Dossier à déposer sur [Démarches en ligne](#)

Référente Intégrité Scientifique (RIS)

- referentintegritescientifique@dauphine.psl.eu

Service Commun de la Documentation (SCD) ou BU

- Pour le PGD et la diffusion des données
- bu-appui-recherche@dauphine.psl.eu

Comité gouvernance, risque et conformité

- Regroupe le RSSI, le DPO et le Fonctionnaire sécurité défense
- [Page ENT](#)



Le PGD pendant la thèse

Quelles questions se poser pendant la thèse ?

Où les données seront-elles stockées et comment seront-elles organisées ?

Quelles sont les procédures de sauvegarde prévues ?

Pour suivre les principes FAIR :

- La méthodologie est-elle assez détaillée pour comprendre comment ont été produites et analysées les données ?
- Comment les données seront-elles décrites ?
- Quel format informatique pour les fichiers ?



Stockage et sauvegarde des données au cours du projet

« Il ne faut pas confondre **stockage sécurisé** (pour assurer la continuité de l'exploitation sur du court terme) et **archivage numérique pérenne** (conservation à moyen et long terme). »

« Toutes les données n'ont pas vocation à être archivées. Les données à archiver doivent présenter une valeur scientifique reconnue par la communauté dont elles proviennent. »

LE STOCKAGE SÉCURISÉ DURANT LE PROJET

GARANTIR LA SÉCURITÉ DES DONNÉES

La règle du 3-2-1 :

- 3 exemplaires : créez deux copies de votre fichier original.
- 2 supports : stockez vos copies sur au moins 2 supports différents.
- 1 copie à distance : stockez au minimum 1 copie hors site.

EXEMPLES

Projet05_v03_2024.pdf sur un ordinateur professionnel

Projet05_v03_2024.pdf sur un serveur institutionnel

Projet05_v03_2024.pdf sur un cloud sécurisé

Solution proposée
à Dauphine (40 Go)



Hors Dauphine
(jusqu'à 90 Go)



DoRANum : « Stockage et archivage » : Fiche pratique. <https://doranum.fr/stockage-archivage/>

Organiser les fichiers et les versions

Définir des règles pour :

Bien nommer ses fichiers

« Dans un ensemble de fichiers ou de documents, des noms « normalisés », cohérents et comportant des éléments discriminants (comme la date de production) permettent également d'appréhender plus facilement le contenu du fichier ou du document avant même de l'ouvrir. Pour cela, il peut par exemple être utile dans le schéma de nommage des fichiers d'inclure des métadonnées importantes (par ex : [Date]_[Run]_[SampleType]. »

Gérer ses différentes versions

« En effet, on peut être amené à devoir revenir sur une version antérieure d'un script si les analyses obtenues ne s'avèrent pas correctes, ou si l'on souhaite modifier la direction que prendrait le projet. »

DATAACC. « Nommage des fichiers et gestion des versions : adopter les bons réflexes ». <https://www.dataacc.org/organiser/nommage-des-fichiers-versioning/>

Opter pour des formats ouverts

Loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique). Article 4 : « On entend par standard ouvert tout protocole de communication, d'interconnexion ou d'échange et tout format de données interopérable et dont les spécifications techniques sont publiques et sans restriction d'accès ni de mise en œuvre. »

Il est recommandé de privilégier les formats ouverts, non propriétaires, dont l'usage est répandu dans la communauté de recherche concernée (formats des fichiers) :

- Cela permet d'éviter l'obsolescence rapide des fichiers.
- Cela garantit de pouvoir accéder aux données dans un temps relativement long via des logiciels open source implémentant le format.
- Cela permet de s'assurer de la réutilisabilité et de la durabilité de ses données.

EPFL FAST GUIDE #4
FILE FORMATS

EPFL LIBRARY
RESEARCH DATA
MANAGEMENT

v3 - 2023
DOI: 10.5681/zenodo.3337829

DEFINITION
A file format is a standard way to encode data as a computer file. It follows a protocol that specifies how bits are used to encode information in a digital storage medium. File formats may be **proprietary** or **open**^[1].

DO NOT FORGET
When listing out the data formats you will be using, make sure to include:
✓ The necessary **software** to view the data (e.g. SPSS v.3; Microsoft Excel 97-2003);
✓ Information about **version control**;
✓ If data are stored in one format during collection and analysis, then converted to another for preservation: list out features that may be lost in data **conversion** such as system-specific labels.

IDEALLY...
... prefer file formats that are:
🔑 **Unencrypted, uncompressed**, commonly used by the research community
🔒 Compliant to an **open, documented** standard: **interoperable** among diverse platforms and applications, fully **published** and available **royalty-free**, fully and independently **implementable** by multiple platform providers on **multiple platforms** without any intellectual property [2]

TYPE OF DATA	APPROPRIATE	ACCEPTABLE	DEPRECATED
Tabular (extensive metadata)	csv - hdf5	txt - html - tex - fastq - por	
Tabular (minimal metadata)	csv - tab - ods - sql - tsv	xml (if appropriate dtd) - xlsx	xls - xlsb
Textual / Presentation	txt - pdf - odt - odm - tex - md - htm - xml - extxyz - odf	pptx - rtf - docx - pdf (with embedded fonts) - eps - ipf	doc - ppt - dvi - ps
Code / Computation	m - r - py - ipynb - rstudio - rmd - NetCDF - aiml	sdd	mat - rdata
Image / Spectroscopy	tif - png - svg - jpeg - fits	jcamp - jpg - jp2 - tif - tiff - pdf - gif - bmp - dm3 - oir - lsm	indd - ait - psd - spc
Audio	flac - wav - ogg - mxl - midi - mel - humdrum	mp3 - aif	
Video	mp4 - mj2 - avi - mkv	ogm - mp4 - WebM	wmv - mov - qt
Geospatial	NetCDF - tabular gis attribute data - shp - shx - dbf - prj - sbx - sbn - postGis - tif - tfw - geoJson	mdb - mif	
3D structures & images	x3d - x3dv - x3db - pdf3D - pov - pdbml	dwg - dxf - pdb	pxp
Other	xml - json - rdf		

WANT MORE?
Check out also these tables:
• fileformats.archiveteam.org/wiki/Scientific_Data_formats
• loc.gov/preservation/resources/rf/TOC.html
• archives.gov/records-mgmt/policy/transfer-guidance-tables.html
• nationalarchives.gov.uk/information-management/manage-information/digital-records-transfer/file-formats-transfer

Credits and sources
[1] en.wikipedia.org/wiki/File_format
[2] guides.library.yale.edu/data-best-practices/format-files

Bibliothèque Contact and info
go.epfl.ch/rdm researchdata@epfl.ch

CC BY-NC-SA

[EPFL Library Research Data Management](#)
[Fastguides](#)



Le PGD à la fin de la thèse

Quelles questions se poser à la fin de la thèse ?

- Les données peuvent-elles être diffusées ?
- Dans quel entrepôt peuvent-elles être déposées ?
- Quelle licence choisir pour diffuser les données ?
- Un archivage, au sens patrimonial du terme est-il prévu ?



Des données diffusables ?



Par défaut, oui (sauf exception listée) !

Anonymiser ou pseudonymiser les données personnelles ?

- **Les données anonymisées** ne sont plus soumises au RGPD et peuvent être utilisées sans restriction.
- L'anonymisation est une opération irréversible, destructrice d'informations.
- Les données peuvent donc être altérées.
- **Les données pseudonymisées** restent soumises au RGPD : les individus ne sont plus identifiables directement.
- La réutilisation des données doit donc se faire sous conditions.
- Contact : dpd@dauphine.psl.eu.



Choisir un entrepôt pour diffuser les données

La préconisation est de privilégier un entrepôt disciplinaire de confiance :

- Pour trouver un entrepôt disciplinaire : [re3data](#) ;
- Liste d'entrepôts disciplinaires de confiance : [Recherche Data Gouv](#) (selon l'analyse réalisée par le Comité pour la science ouverte).

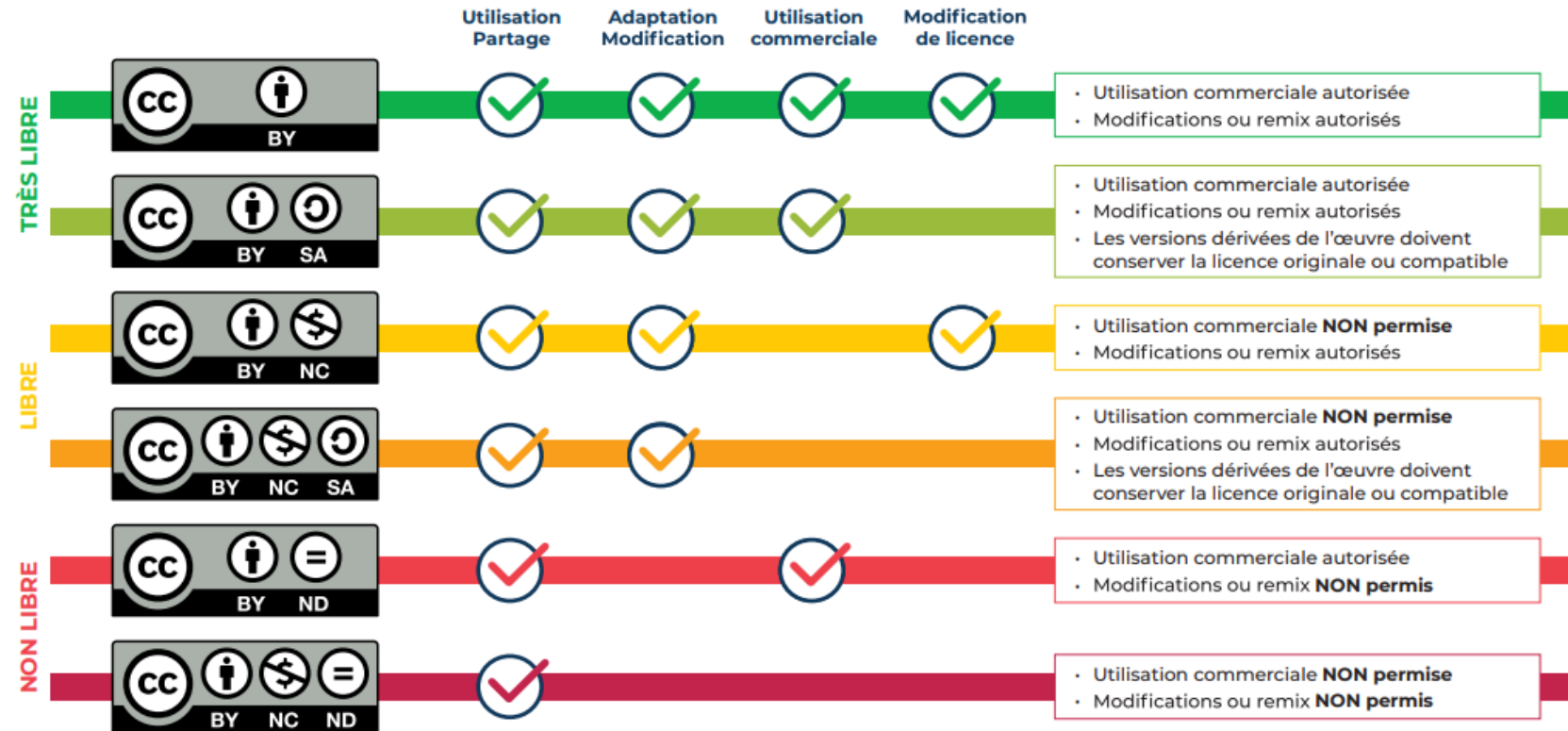
En l'absence d'un entrepôt disciplinaire :

- [Zenodo](#) : entrepôt à vocation généraliste
- [Recherche Data Gouv](#) : une [collection pour les jeux de données dauphinois](#) a été créée fin 2024.



Les licences Creative Commons (CC)

Schéma des licences Creative Commons



BY (attribution to the creator)

ATTRIBUTION

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

L'auteur doit être cité



SA (Share Alike)

PARTAGE DANS LES MÊMES CONDITIONS

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Partage sous licence compatible



NC (Non Commercial)

POUR USAGE NON COMMERCIAL

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Pour usage non commercial



ND (No Derivative works)

PAS DE MODIFICATION

Vous pouvez retenir, réutiliser et redistribuer.

Aucune modification permise

Source : Les Licences. (n.d.). *Fabrique REL*.
<https://fabriquerel.org/licences/>



L'auteur de l'œuvre doit être cité en toutes circonstances pour l'ensemble de ces licences.

Cette œuvre est mise à disposition selon les conditions de la Licence Creative Commons Attribution 4.0 International.



Archiver les données de la recherche

« L'archivage numérique pérenne consiste à conserver les fichiers de données, les logiciels et leurs métadonnées :

- Dans leur aspect physique et intellectuel,
- Sur le **moyen et long terme** (plus de 30 ans)
- De manière à ce qu'ils soient en permanence **accessibles et compréhensibles** »

« Le **CINES** (Centre Informatique National de l'Enseignement Supérieur) a été mandaté par le Ministère de l'Enseignement supérieur et de la Recherche pour mettre en place une solution d'archivage électronique du patrimoine scientifique pour toute la communauté ESR (Enseignement Supérieur et Recherche). »

Bonnes pratiques	Stockage	Archivage pérenne
Sécuriser	<ul style="list-style-type: none"> • Suivre la règle du 3-2-1 • Avoir des supports fiables (serveur et cloud institutionnels) • Chiffrement des données • Antivirus, mot de passe 	<ul style="list-style-type: none"> • Sécurité matérielle et logicielle assurée par l'opérateur national CINES
Rendre accessible	<ul style="list-style-type: none"> • Utiliser une plateforme collaborative de travail accessible aux partenaires du projet • Versioning des fichiers 	<ul style="list-style-type: none"> • Fichiers encore accessibles après 30 ans • Niveau d'accès sur mesure • Accès sur demande à l'institution déposante
Préserver l'intégrité physique	<ul style="list-style-type: none"> • Éviter la perte de données sur des supports nomades et fragiles • Suivre la règle du 3-2-1 	<ul style="list-style-type: none"> • Se protéger de la détérioration des supports • S'assurer que l'information contenue sur le média n'est pas altérée
Préserver l'intelligibilité du contenu	<ul style="list-style-type: none"> • Bien nommer ses fichiers • Documenter les données et renseigner les métadonnées au fur et à mesure du projet 	<ul style="list-style-type: none"> • Contextualiser l'information • Garantir une bonne compréhension des données sur le long terme
Préserver la lisibilité	<ul style="list-style-type: none"> • Privilégier les logiciels et formats ouverts • Planifier les opérations de migration des supports 	<ul style="list-style-type: none"> • Surveillance de l'obsolescence, des standards et mise en place de dispositifs de migration matérielle et logicielle

Contactez le service Archives de Dauphine :
contact.services.archives@dauphine.psl.eu

DoRANum : « Stockage et archivage » : Fiche pratique.
 DOI : [10.13143/0c4b-2743](https://doi.org/10.13143/0c4b-2743)



Remplir son PGD dans DMP OPIDoR

Remplir son PGD dans DMP OPIDoR : récapitulatif

Informations sur mes données	Champ(s) à remplir dans DMP OPIDoR	Recommandations
Réutilisation de données existantes	1.2 Est-ce que des données existantes seront réutilisées ?	Bien vérifier les réutilisations possibles des données.
Création / Collecte de nouvelles données	1.3 Comment seront produites/collectées les nouvelles données ?	Documenter le plus précisément possible (logiciels, procédures, etc.) pour permettre la réutilisation.
Droit d'auteur / propriété des données	3.1 Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?	Si financement à 50% et plus par des fonds publics = données publiques (avec exceptions). Si financement à moins de 50% par des fonds publics = vérifier les contrats.
Non divulgation des données	6.1 Comment les données seront-elles partagées ?	Indiquer les raisons de la non-diffusion des données.

Remplir son PGD dans DMP OPIDoR : récapitulatif

Informations sur mes données	Champ(s) à remplir dans DMP OPIDoR	Recommandations
Données personnelles / RGPD	3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?	Contacter le DPO de Dauphine.
Anonymisation des données	3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?	<ul style="list-style-type: none"> - Différencier anonymisation et pseudonymisation. - Indiquer la méthodologie utilisée.
Financier de la recherche	Onglet « Informations générales » du PGD	Permet de récupérer automatiquement les informations du projet.
Éthique et intégrité scientifique	3.3 Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?	Si besoin, contacter les référents ou comités dédiés.
Stockage et sauvegarde des données au cours du projet	5.1 Comment les données seront-elles stockées et sauvegardées tout au long du projet ?	Bien détailler le stockage (usage courant) et la sauvegarde (en cas de perte notamment).

Remplir son PGD dans DMP OPIDoR : récapitulatif

Informations sur mes données	Champ(s) à remplir dans DMP OPIDoR	Recommandations
Organisation des fichiers et des données au cours du projet	2.1 Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?	Important pour visualiser les différentes versions des données (brutes, traitées, à diffuser).
Format des données	1.1 Description générale du produit de recherche	Utiliser de préférence des formats ouverts.
Choix d'un entrepôt de diffusion pour les données	6.1 Comment les données seront-elles partagées ?	Une liste déroulante d'entrepôts est proposée.
Choix d'une licence pour les données diffusées	3.1 Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?	Une licence ouverte indique les droits de réutilisation accordés a priori.
Archivage des données	6.2 Comment les données seront-elles conservées à long terme ?	Contactez le service des archives.



**Valoriser le travail
autour des données**

Comment valoriser son implication dans la gestion des données ?

Lier les publications et les données

- [Partager les données liées aux publications scientifiques : guide pour les chercheurs \(2022\)](#) ;
- Utiliser un identifiant unique pérenne pour citer vos données dans vos publications.

Partager son PGD

- Le rendre public dans [DMP OPIDoR](#) ;
- Le [déposer dans HAL](#).

Écrire un Data Paper

- « Article scientifique évalué par les pairs et citable. Il décrit un jeu de données (data, data set), la méthode ayant permis de l’obtenir et le potentiel de réutilisation de ce jeu. » [Cirad, 2022](#) ;
- Valorisation des [résultats dits négatifs](#) ;
- Cours en ligne « [Rédiger et publier un Data Paper](#) » (permet d’obtenir un Open Badge).

Participer à un prix

- Prix Science Ouverte 2026 : [Thèse, Données, Logiciel libre](#).





Pour aller plus loin

Pour en apprendre plus sur les données



Publications du Comité pour la Science Ouverte

Comment appliquer ces principes dans son quotidien

↳ [Toutes les publications](#)

↳ [Guides Passeport](#)



DoRANum

Membre de l'écosystème Recherche Data Gouv, centre de ressources consacré à la formation

↳ [DoRANum](#)



Fiches pratiques IPOS

Fiches pratiques synthétiques et locales rédigées par la BU, autour des données et de la Science Ouverte en général

↳ [Fiches pratiques IPOS](#)

Pour de l'aide sur vos recherches documentaires et l'utilisation des ressources



QoRAiL

Service de question-réponse en ligne

Une réponse en 72h

↳ bu.dauphine.psl.eu/qorail



Coaching

Un rendez-vous personnalisé avec un bibliothécaire, seul ou en petit groupe, en présentiel ou en ligne via Teams

↳ [Formulaire de demande](#)



Mail

↳ bu-appui-recherche@dauphine.psl.eu

Pour des questions pratiques sur les services de la BU



Mail

↳ bu@dauphine.psl.eu



Bureaux d'information

Les bibliothécaires sont disponibles aux différents points d'accueil des BU



Site web et réseaux sociaux

↳ bu.dauphine.psl.eu

↳ [Instagram](#), [YouTube](#), [LinkedIn](#), [Bluesky](#)

Merci pour votre participation !

Renaud Delemontez-Sage, Aurélie Gono, Fedor Reshetnikov

bu-appui-recherche@dauphine.psl.eu

Bibliothèque de l'université Paris Dauphine-PSL



UNIVERSITÉ PARIS DAUPHINE - PSL

Place du Maréchal de Lattre de Tassigny – 75775 Paris cedex 16