

Gérer ses données de recherche en sciences humaines et sociales

Formation doctorale ED 450

10 février 2026

ENS-PSL

Elifsu Sabuncu (CNRS) & Lina Sbeih (ENS)

Objectifs de la séance

À l'issue de ce cours, vous serez en mesure de :

- connaître les enjeux autour des données de la recherche
- identifier, organiser et documenter vos données
- répondre aux exigences institutionnelles

Plan

1. Introduction : définitions et cadre
2. Données de la recherche : droit et éthique
3. Le cycle de vie des données de la recherche
4. Organiser et documenter ses données au quotidien
5. Mini-atelier : Organiser ses données en 10 minutes
6. Ouvrir et valoriser ses données
7. Le plan de gestion des données (PGD)
8. Conclusion, ressources...et action !

1. Introduction : définitions et cadre

1.1.Diversité des données en SHS

1.2 Multiplicité des enjeux

définition ?

Données de la recherche : définitions

- archivistes
- OCDE

Les « données de la recherche^[1] » sont, selon les archivistes de l'ESR :

« l'ensemble des informations, spécimens et matériaux produits, recueillis et documentés par les chercheurs, et qui sont collectées et exploitées à des fins de recherche et de preuve par les chercheurs et leurs équipes. »

A ce titre,

« elles constituent, sur le long terme, une partie des archives de la recherche. »

« les archives de la recherche englobent l'ensemble des documents et données produits ou reçus dans le cadre du processus de recherche ».

Autrement dit, ces données de la recherche couvrent donc :

- l'activité de recherche au sein des laboratoires et des équipes de recherche
- l'administration et le financement des projets de recherche, et les fonctions supports de la recherche (archives, documentation, informatique).

« les données de la recherche sont, aujourd'hui, en grande partie électroniques, mais peuvent exister aussi sur d'autres supports. »

Source : La Section Aurore de l'AAF : le réseau professionnel dédié aux archives de l'enseignement et de la recherche

Selon l'OCDE :

« Les données de la recherche sont définies 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. »

Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics (rapport 2007)

https://www.ouvrirlascience.fr/wp-content/uploads/2018/11/Principes-lignes-directrices-de-l%E2%80%99OCDE-pour-l%E2%80%99acc%C3%A8s-aux-donn%C3%A9es_38500823.pdf

1.1. Diversité des données en SHS

(Une rapide et incomplète vue d'ensemble)

des exemples ?

Plusieurs distinctions possibles :

- Données produites/données collectées
- Données qualitatives/quantitatives
- Données sensibles
- etc.

1.2. Multiplicité des enjeux

**à votre avis, quels enjeux,
pourquoi s'occuper de ses
données?**

Enjeux et motivations

- S'y retrouver...
- Preuve (juridique) et valorisation du travail (doctoral)
- Continuité des recherches au sein des laboratoires, malgré les changements de personnes
- Bonnes pratiques scientifiques : transparence, traçabilité, reproductibilité, intégrité
- Attentes institutionnelles : votre ED vous y oblige souvent ; pour le chercheur/labo -> ses tutelles (CNRS, etc.)
- Patrimoine, mémoire de la recherche, histoire des sciences
- Exigences des financeurs (ANR, Europe)
- Développement de la science ouverte + open data -> Loi Lemaire 2016

Loi Lemaire : ce qu'il faut en retenir

En France, la Loi Lemaire (*La Loi pour une République numérique, 2016*), acte le principe d'ouverture **par défaut** des **données publiques (mouvement mondial de l'open data)**. Et comme les données produites dans le cadre d'une recherche financée (au moins pour moitié) par des fonds publics sont assimilées à des données publiques, elles sont concernées par la loi.

Ce qui veut dire que :

- 1/ les **données de la recherche ne sont pas la propriété d'une administration ni d'un éditeur (=la question des publications scientifiques)**,
- 2/ leur **réutilisation est libre** à partir du moment où elles ont été rendues publiques, sauf si elles sont soumises à un droit spécifique ou une réglementation particulière

Conséquences :

1/ **Publis** -> Les auteurs ont le **droit de déposer leur version d'article (preprint et postprint) dans des archives ouvertes** (HAL, etc.) après un délai d'embargo (6 mois en sciences, 12 mois en SHS) -> l'éditeur ne peut pas restreindre ni la publication des résultats, ni la réutilisation des données associées aux publications.

2/ Données FAIR

3/ développement **d'infrastructures publiques** permettant d'organiser et de retrouver ces données : dépôts nationaux et disciplinaires -> huma-num qui était là depuis longtemps + création d'<https://recherche.data.gouv.fr/>, etc.

FAIR vs OPEN data

FAIR data $\neq$ Open data

respecting any ethical, legal or contractual restrictions:

- well-defined access protocols
- permission request possible

Completely open data can be used, shared and built-on by anyone, anywhere and for any purpose

FAIR data can be 'not open'

- a data set is findable, reusable, interoperable
- but only accessible within a closed group

Open Data can be 'not FAIR'

- undocumented data dump in an uncurated page
- not findable, reusable, interoperable

FAIR vs OPEN data

Point commun : promotion de la transparence, de l'innovation et de la collaboration dans le domaine scientifique

Différences :

- OPEN -> publication sans restriction
- FAIR -> garantissent que ces données sont organisées et documentées de manière à être facilement accessibles et utilisables par d'autres

La base : bien gérer ses données en tant que bonne pratique, pour des recherches de bonne qualité et efficace. Ensuite, vous pouvez aller encore plus loin et rendre vos données FAIR en utilisant des (méta)données standardisées et lisibles par machine, des identifiants permanents, des licences d'utilisation des données et des formats ouverts.

Mieux : FAIRiser vos données va vous aider à aller vers (un vrai) OPEN.

Attention, les données OPEN ne sont pas toujours FAIR, ni même correctement gérées : diffusées des données ouvertes mais techniquement inutilisables, non structurées, non documentées, etc.

à retenir : Tous les bailleurs de fonds recommandent ou exigent que les données de recherche soient FAIR et aussi ouvertes que possible.

2. Données de la recherche : droit et éthique

- 2.1. Archives publiques
- 2.2. Données à caractère personnel
- 2.3. Propriété intellectuelle
- 2.4. Intégrité scientifique

2.1. Archives publiques

Livre 2 dédié aux archives du code du patrimoine :

« Les archives sont l'ensemble des **documents et des données**, quels que soient leur **date, leur forme, leur support matériel, produits ou reçus par toute personne physique ou morale et par tout organisme public ou privé dans l'exercice de leur activité.** » (Code du patrimoine L.211-1).

PSL étant un établissement public, nos archives sont publiques. Elles sont donc imprescriptibles et inaliénables, soumises à des règles de gestion spécifiques (destruction, versement des archives historiques dans des services d'archives publiques).

Les archives sont donc la propriété de votre institution si vous êtes payés. Vous avez la responsabilité de les gérer correctement et ne pouvez pas en faire ce que vous voulez (par exemple les détruire).

Régime de communication des archives ou délais de communicabilité. Les archives publiques sont, communicables de plein droit, sauf si elles contiennent les secrets protégés par la loi

Il existe des services d'archives et des archivistes dans la plupart des établissements de PSL dont EPHE, Dauphine et ENS.

2.2. Données à caractère personnel

Depuis 1978, le droit français s'est doté de la loi relative à l'informatique, aux fichiers et aux libertés et de la CNIL.

L'objectif du règlement général à la protection des données (RGPD), ou règlement relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation des données (UE 2016/679) de 2018, est, d'harmoniser la réglementation de l'UE et selon son article 1, de « protéger les libertés et droits fondamentaux des personnes physiques » en luttant contre le contrôle de l'utilisation des données à caractère personnel notamment par les grands acteurs du web.

2.3. Propriété intellectuelle

La propriété intellectuelle regroupe

- la propriété industrielle et
- la propriété littéraire et artistique (Code de la propriété intellectuelle).

La propriété industrielle protège et valorise les inventions grâce notamment au dépôt de brevet.

L'œuvre de l'esprit est protégée par le droit d'auteur du fait même de son existence.

- Elle doit être originale et formalisée.
- S'y appliquent :
 - des droits moraux (paternité, respect et intégrité de l'œuvre...) qui sont perpétuels, inaliénables et imprescriptibles et
 - des droits patrimoniaux c'est à dire des droits d'exploitations donc de reproduction (en France, une œuvre entre dans le domaine public 70 ans après la mort de son auteur).

2.4 Intégrité scientifique

D'après l'office français de l'intégrité scientifique : L'intégrité scientifique renvoie à l'ensemble des règles et des valeurs qui doivent régir les activités de recherche pour en garantir le caractère honnête et rigoureux. Selon le code de la recherche ([article L. 211-2](#)), elle « contribue à garantir l'impartialité des recherches et l'objectivité de leurs résultats ».

Principes essentiels : **la fiabilité, le respect, l'honnêteté, la responsabilité**. L'article 6 du [Décret n° 2021-1572 du 3 décembre 2021](#) relatif au **respect des exigences de l'intégrité scientifique dans les établissements de recherche** indique : « Les établissements publics [...] définissent une politique de **conservation**, de **communication** et de **réutilisation** des **résultats bruts** des travaux scientifiques menés en son sein. À cet effet, ils veillent à la mise en œuvre par leur personnel de **plans de gestion de données** et contribuent aux infrastructures qui permettent la conservation, la communication et la réutilisation **des données et des codes sources**. »

Il existe un référent intégrité scientifique dans les établissements

des questions ?

3. Le cycle de vie des données de la recherche

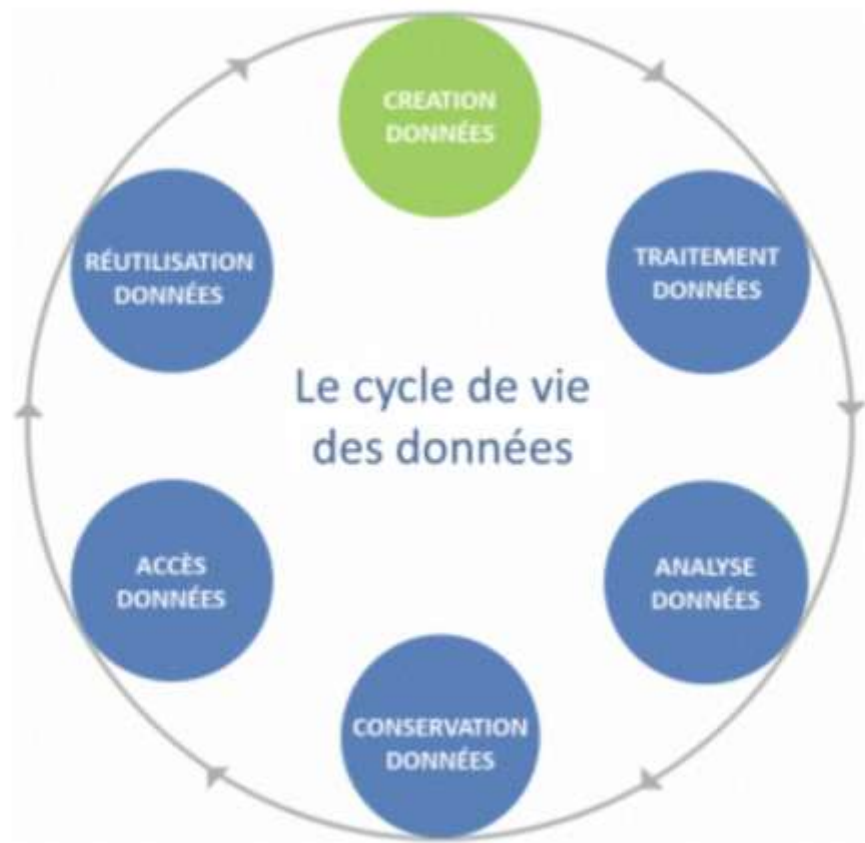
3.1. Les étapes du cycle de vie des données

3.2. Zoom sur : stockage - sauvegarde - pérennisation

3.3. “Et moi dans tout ça ?”

- Rôle et responsabilité des doctorants quant aux données
- Moments critiques de la thèse

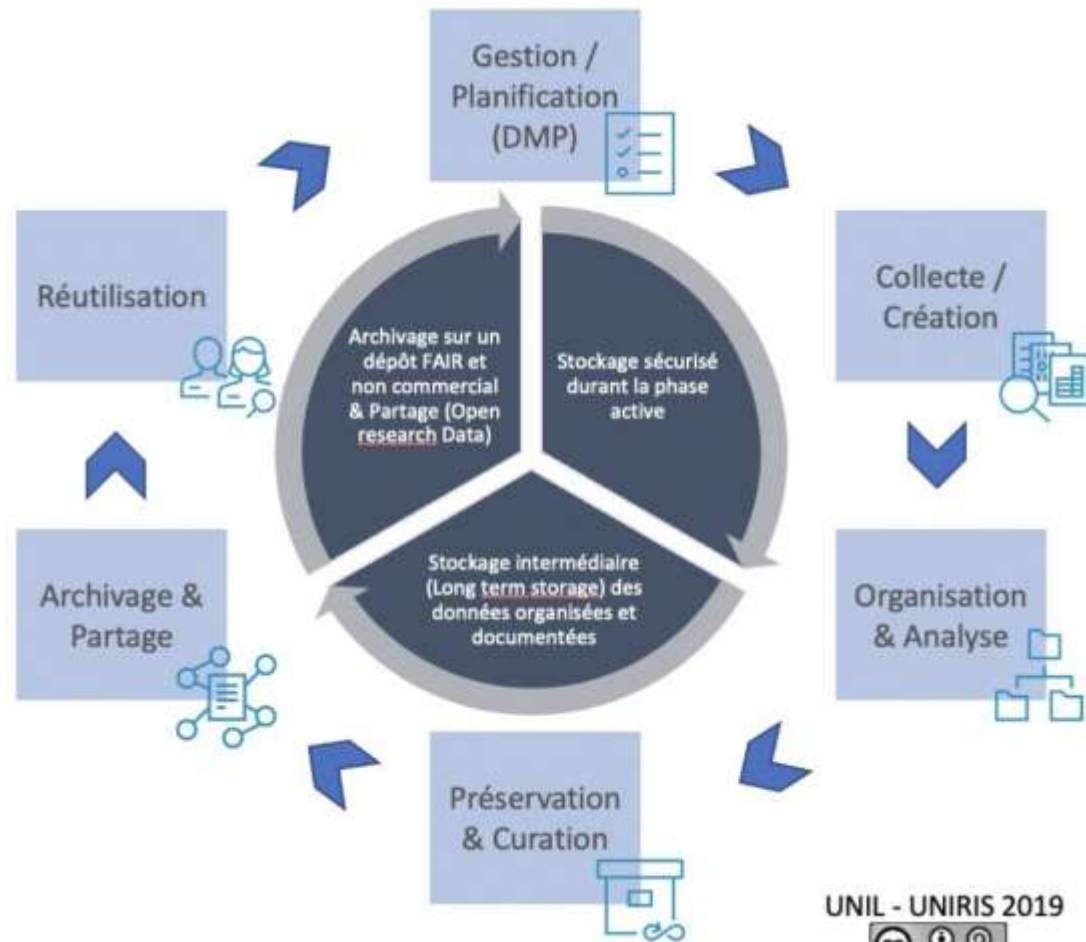
3.1 Les étapes du cycle de vie des données



Cycle de vie des données de la recherche (adapté de UK Data Archive)







UNIL - UNIRIS 2019



3.2 stockage - sauvegarde - pérennisation

3.3. “Et moi dans tout ça ?”

?

Rôle et responsabilité des doctorants quant aux données

Ce que vous maîtrisez :

- L'organisation quotidienne, la structuration, nommage, etc.
- Le choix de formats (ouverts, standards, pérennes)
- La documentation (décrire les données, contexte d'obtention, etc.)
- La conservation : le stockage sécurisé, contrôler l'accès, protéger la confidentialité
- La sauvegarde pour la transmission
- Les décisions de diffusion ou de restriction
- (en partant) Laisser une copie documentée, avec un plan de classement (charte de nommage)

Moments critiques de la thèse

Certaines phases sont plus à risque (perte partielle ou totale des données, perte des infos de contexte, non conformité des formats...) :

- le terrain -> on est loin du labo, parfois isolé
- la transcription et l'analyse -> il ne faut pas perdre ses données, et bien les organiser pour s'y retrouver
- les changements d'ordinateur (je ne parle pas de perte ou vol de l'ordinateur portable)
- la rédaction (de la thèse et des publications) -> comment citer ses données
- l'après-soutenance -> qu'est-ce qu'on prend, qu'est-ce qu'on laisse ?

4. Organiser et documenter ses données au quotidien

- 4.1. Nommer ses fichiers
- 4.2. Structurer ses dossiers
- 4.3 Versions et sauvegardes
- 4.4 Documenter ses données

4.1 Nommer ses fichiers

Petit jeu

EntretienMarieFinal(2).docx

2024-03-12_ENTR_MarieDupont_v1.txt

TableauDonnees.xlsx

CV_new.doc

2024-07-01_Stats_Sondage_AcadFR_v2.csv

4.1 **Bien** nommer ses fichiers...pourquoi ?

4.1 Nommer ses fichiers numériques : les bonnes pratiques

LES BONNES PRATIQUES DE NOMMAGE

- Les intitulés de fichiers doivent être succincts et précis et ne pas dépasser **31 caractères** au maximum, extension comprise¹. L'ensemble (nom + chemin du fichier) est limité à 256 caractères dans le système d'exploitation Microsoft.
- **Ne pas utiliser les signes diacritiques** : pas d'accent ni de tréma (é, è, ê, à, ä, ù), pas de cédille (ç)
- **Ne pas utiliser les caractères spéciaux** et les espaces vides (sauf underscore « _ ») comme : ; , . ! ? ' / \ : * " < > % () @ & etc.
- **Ne pas utiliser de mots vides** : le, la, les, un, une, des, et, ou etc
- **Ne pas utiliser de dénominations vagues**, comme : « divers » « autres », « à »
- **Ne pas utiliser le nom de l'agent créateur** ou gestionnaire du dossier et du fichier. Exemples : « martin », « dupont ».

4.1 Bien nommer ses fichiers numériques : les éléments

- nom du projet
- sujet
- type de document
- date
- version
- auteur

4.1 Bien nommer ses fichiers : exemples

Exemple de bonne pratique de nommage

codeuniv_codeprojet_sujet_typededocument_annee_mois_jour_version

DGDFA_SIFAC_reunion_CR_2020_01_15_V3F

UP_Arch_fiche_pratique_nommer_fichiers_dossiers_V2F.pdf

4.2 Structurer ses dossiers

4.2 Structurer ses dossiers : pourquoi ?

4.2 Structurer ses dossiers : exemples

Exemple d'arborescence pour un métier (fusion)

- 001_PILOTAGE
 - 01_COPIL_ADM
 - 02_COPIL_METIER
 - 03_INDICATEURS
 - 04_ORGANISATION_CIBLE
 - 05_METHODOLOGIE
- 002_CHANTIER_X
 - 01_FICHE_CHANTIER
 - 02_SUIVI_DE_PROJET
 - 03_LIVRABLE_1
 - 04_LIVRABLE_2
- 003_CHANTIER_Y
- 004_CHANTIER_Z

Modèle de plan de classement

1. Principale Fonction n°1 [Administration]

1.1. Activité n°1 [Correspondance]

1.2. Activité n°2 [Réunions]

1.2.1. Processus n°1 [Réunions de service]

- Document X [PV de réunion]
- Document Y [Convocations]

1.3. Activité n°2 [Réunions extérieures]

- Document X [PV de réunion extérieure]

2. Principale Fonction n°2 [Gestion des RH]

2.1. Activités n°1 [Recrutement]

Projet_SHS_These/

|

|— 01_Protocoles/

| |— Grille_entretien.pdf

| |— Consentement_informe.docx

||

|— 02_Donnees_brutes/

| |— Audio/

| |— NotesTerrain/

|

|— 03_Transcriptions/

|

|— 04_Analyses/

| |— Codes/

| |— Tableaux/

|

|— 05_Diffusion/

| |— Article_preprint.pdf

| |— Article_postprint.pdf

|

|— README.md // metadonnees.csv

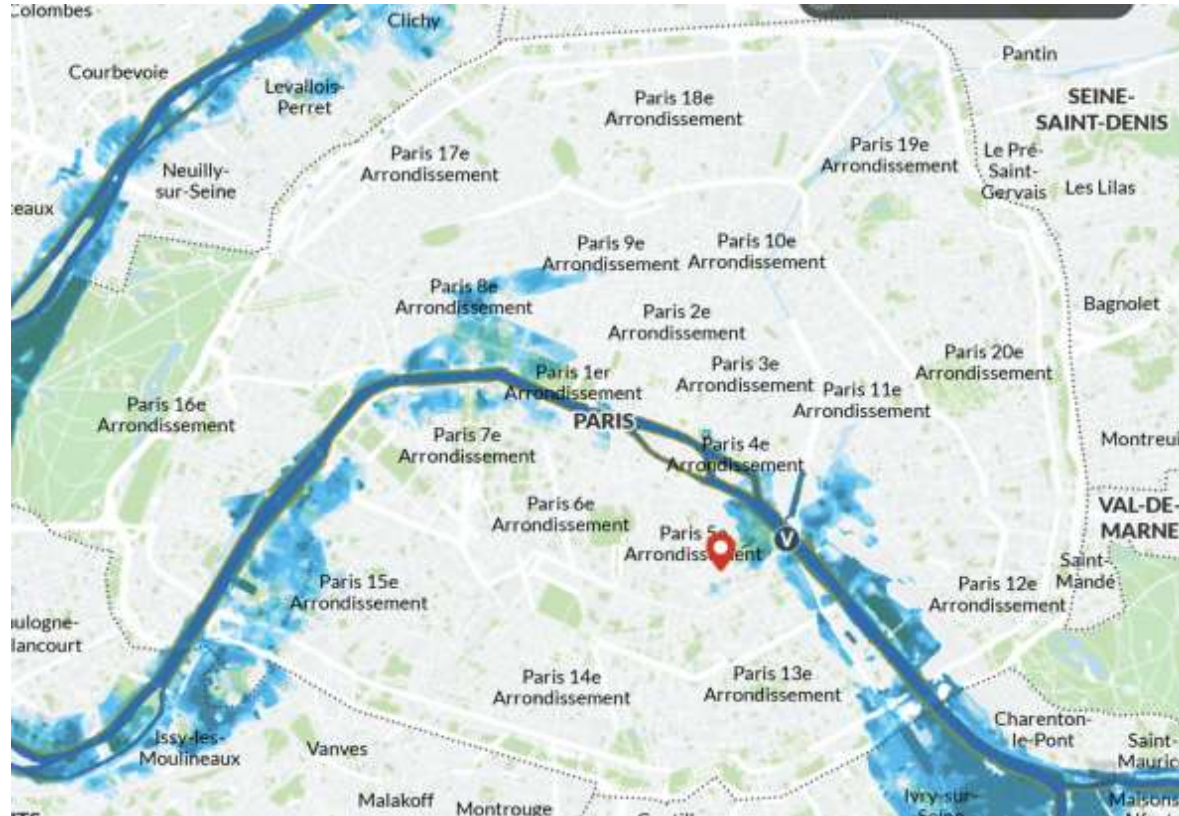
4.3. Versions et sauvegardes

4.3. Versions et sauvegardes :

la règle des 3-2-1

?

4.3. Versions et sauvegardes : la règle des 3-2-1



4.3. Versions et sauvegarde la règle des 3-2-1

ZONES INONDÉES POTENTIELLES - ZIP

① Faites varier indépendamment le niveau des cours d'eau visibles à l'écran



Identifier les impacts potentiels d'une crue par débordement des grands cours d'eaux pour sensibiliser la population et aider à s'y préparer.

PRÉFET
DE LA RÉGION
D'ÎLE-DE-FRANCE

Direction régionale et interdépartementale
de l'environnement, de l'aménagement
et des transports d'Île-de-France

L'INSTITUT
PARIS
REGION



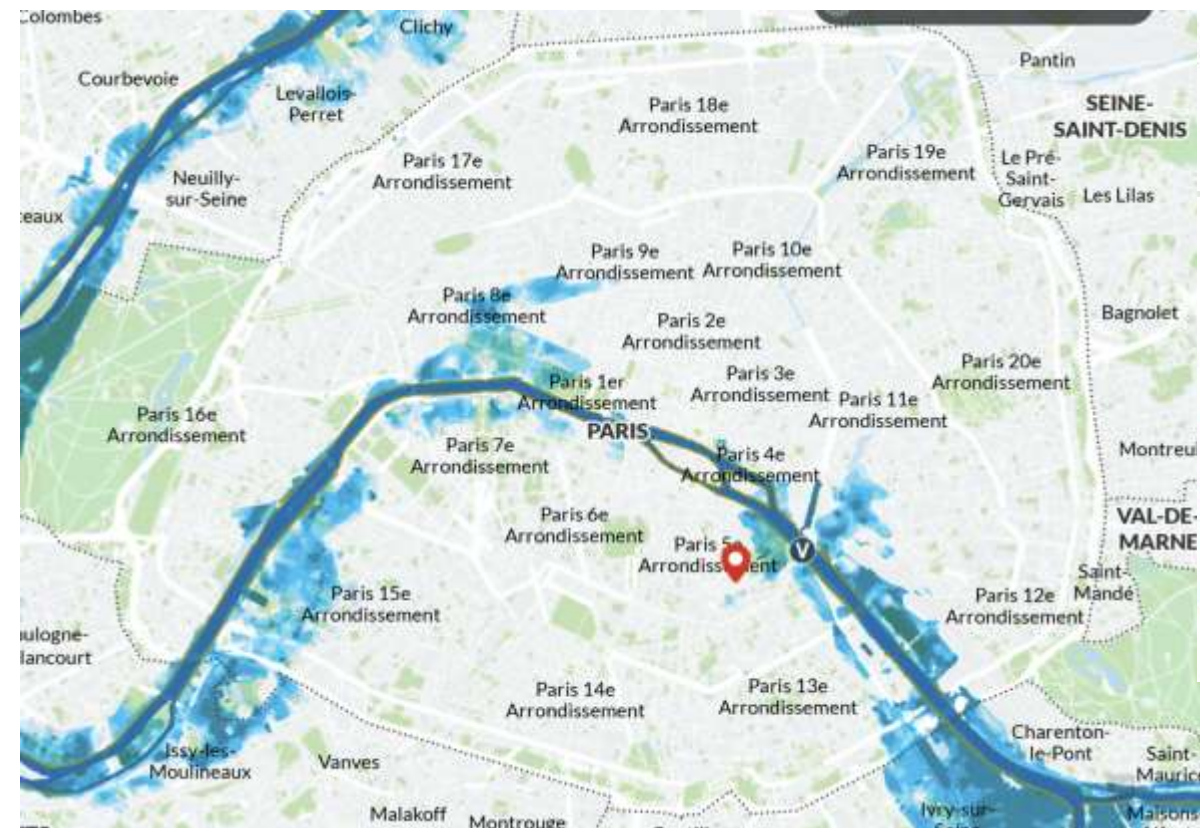
HAUTEUR D'EAU MAXIMALE ESTIMÉE : ENTRE 0,5 MÈTRE ET 1 ...

VIGICRUES

Station : Paris-Austerlitz (Seine à Paris)

Hauteur d'eau à l'échelle de référence de la station : 8.12 m

Crue de référence : 28 janvier 1910 (8,62 m)



4.3. Versions et sauvegardes : la règle des 3-2-1:

3 copies
2 supports
1 hors site

4.4 Documenter ses données

4.4 Documenter ses données : pourquoi ?

Exemples de fichiers de métadonnées

Ajouter un fichier CS : un tableau pour décrire vos données

Colonnes : (standard dont Dublin Core) :

titre,

auteur/créateur/provenance/producteur,

type de document,

sujet,

identifiant/numéro,

date,

format,

volumétrie,

langue,

secrets protégés/données à caractère personnel,

durée de conservation/cert final (conserver ou détruire) en prenant en compte l'intérêt des données

1. Title (Titre)

2. Creator (Créateur)

3. Subject (Sujet)

4. Description (Description)

5. Publisher (Éditeur)

6. Contributor (Contributeur)

7. Date (Date)

8. Type (Type)

9. Format (Format)

10. Identifier (Identifiant de la ressource)

11. Source (Source)

12. Language (Langue)

13. Relation (Relation)

14. Coverage (Couverture)

15. Rights (Gestion des droits)

Exemple : Readme.txt

Readme.txt

Projet/sujet : Étudiants et engagement citoyen

Date : mars-décembre 2026

Chercheur : Paul Martin (Unité, Université (PSL))

Données : 12 entretiens audio (.ogg) + transcriptions (en .txt)

Lieu de stockage/Serveur/Arborescence : /Donnees brutes /Transcriptions /Analyses

Licence : CC-BY-NC-SA, données anonymisées

des questions ?

5. Mini-atelier :

Organiser ses données en 10 minutes



Étape 1 : identifier ses données (et les documenter !)

Prenez une feuille ou votre ordinateur. Notez simplement les types de données que vous produisez ou utilisez. Pas besoin d'être exhaustif. Posez-vous surtout la question des formats et des données sensibles.

En 3 minutes :

- quels types de données produisez-vous ?
- sous quels formats ?
- lesquelles sont sensibles ?

Étape 2 : créer une arborescence

*Imaginez que quelqu'un (vous ?) ouvre votre dossier dans 5 ans. Est-ce que la structure est compréhensible ?
Travail, sources, analyses, résultats : restez simple.*

En 4 minutes :

- proposer une arborescence simple
- distinguer travail / sources / résultats
- penser à l'après-thèse

Étape 3 : établir une règle de nommage

Choisissez UNE règle que vous pourrez tenir sur la durée. Une règle imparfaite mais stable est meilleure qu'une règle idéale jamais appliquée.

En 3 minutes :

- définir une règle unique
- applicable à tous les fichiers
- compréhensible par d'autres (dont le futur vous)

Étape 4 : sauvegarder

Si votre ordinateur disparaît ce soir, que reste-t-il ? Une sauvegarde n'est une sauvegarde que si elle est à jour et distincte.

En 2 minutes :

- où sont stockées vos données ?
- en combien de copies ?
- que se passe-t-il en cas de perte ?

Atelier : débrief

6. Ouvrir et valoriser ses données

6.1. Pourquoi publier ses données ?

6.2 Où déposer ses données ?

6.3 Comment citer des données ?

6.1 Pourquoi publier ses données ?

Partager/publier/ouvrir ses données, c'est aller encore plus loin que la "simple" bonne gestion (pour soi/son labo/sa discipline/son financeur etc.), c'est aller vers la science ouverte (la rencontre de l'open data et de la recherche) :

- favoriser la transparence/traçabilité/reproductibilité/intégrité
- permettre la citation (une citation de données doit pointer vers un objet durable, donc avec identifiant pérenne, donc publié dans un entrepôt)
- permettre la réutilisation
- accroître la visibilité/valoriser

6.2 Où déposer ses données ?

Dans un entrepôt structuré pour les accueillir :

2 principales pistes :

- entrepôts disciplinaires (astro, biologie, SHS, archéologie, etc.) -> NAKALA d'huma-num pour les SHS
- entrepôts institutionnels souvent sur Recherche Data Gouv Recherche Data Gouv, Inauguré en 2022 par la ministre de l'ESR, est *“l'écosystème national dédié à la gestion, au partage et à l'ouverture des données de la recherche française”* ou encore *“une [plateforme centrale](#) pour la préservation, la publication et la découverte des jeux de données”*
 - PSL -> <https://entrepot.recherche.data.gouv.fr/dataverse/psl>

6.3 Comment citer des données ?

des questions ?

7. Le plan de gestion des données (PGD)

7.1 Le PGD : définitions

7.2 Le PGD : le contenu et les rubriques

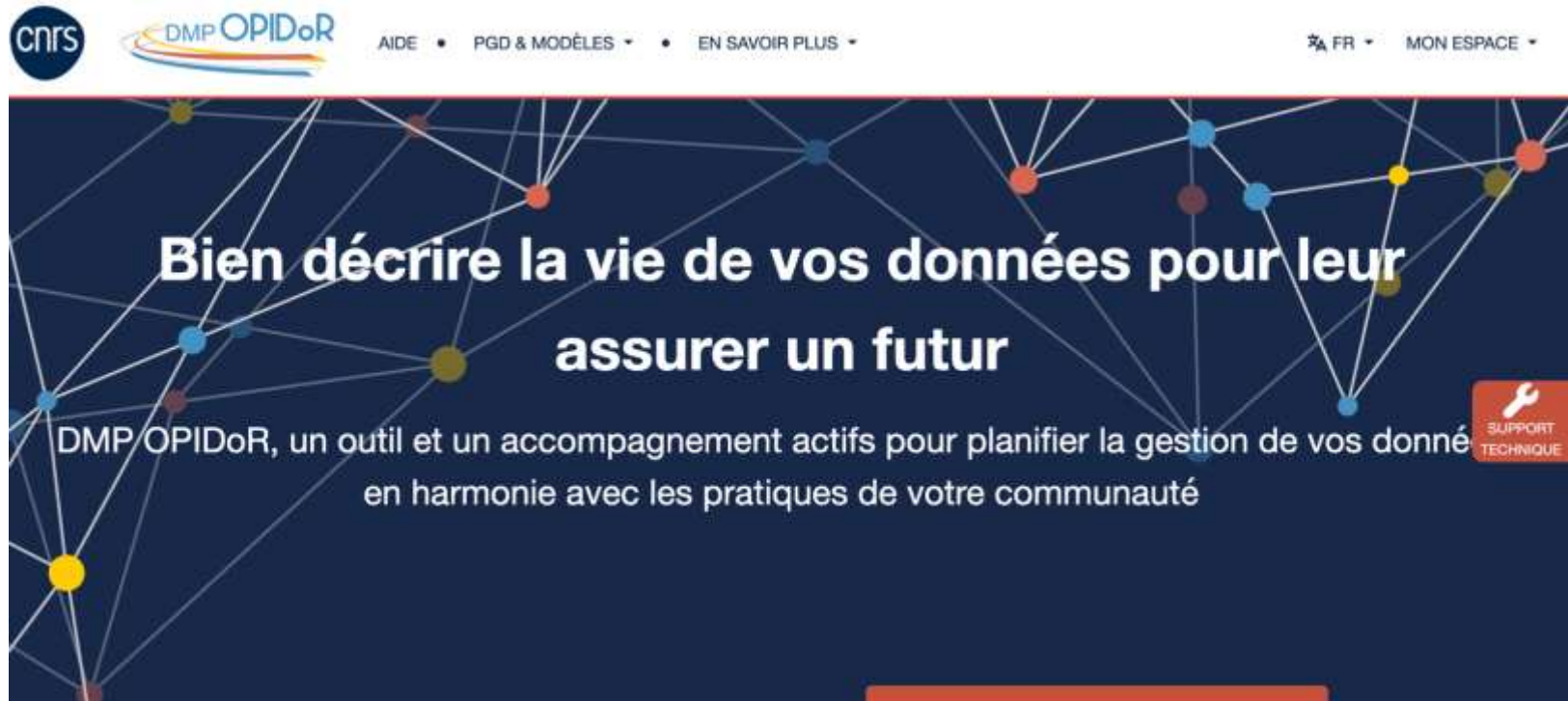
7.3 Le PGD : les outils

7.1 Le PGD : définition

7.2 Le PGD : le contenu et les rubriques

- ØPrésentation du projet
- ØPrésentation des données
- ØAspects juridiques, éthiques, légaux
- ØStockage des données durant le projet
- ØDiffusion et partage à l'issu du projet / Archivage sur le long terme /
- ØResponsabilités

7.3 Le PGD : les outils



The image shows the top section of the DMP OPIDoR website. At the top left is the CNRS logo. Next to it is the DMP OPIDoR logo, which consists of the text 'DMP OPIDoR' with a stylized orange and blue swoosh underneath. To the right of the logo are three navigation links: 'AIDE', 'PGD & MODÈLES', and 'EN SAVOIR PLUS', each followed by a small downward arrow. Further right are two more links: 'FR' (with a small flag icon) and 'MON ESPACE', also with downward arrows. Below the navigation bar is a large dark blue banner with a network diagram of white lines and colored dots (blue, yellow, orange, green). The main text on the banner is in white: 'Bien décrire la vie de vos données pour leur assurer un futur'. Below this, in a smaller white font, is the text: 'DMP OPIDoR, un outil et un accompagnement actifs pour planifier la gestion de vos données en harmonie avec les pratiques de votre communauté'. In the bottom right corner of the banner is a red square button with a white wrench icon and the text 'SUPPORT TECHNIQUE'.

cnrs

DMP OPIDoR

AIDE • PGD & MODÈLES • EN SAVOIR PLUS

FR MON ESPACE

Bien décrire la vie de vos données pour leur assurer un futur

DMP OPIDoR, un outil et un accompagnement actifs pour planifier la gestion de vos données en harmonie avec les pratiques de votre communauté

SUPPORT TECHNIQUE

<https://dmp.opidor.fr/>

8. Conclusion, ressources...et action !

Par où commencer concrètement ?

1 -> École doctorale !!

puis :

- Huma-Num (ShareDocs)
- Labo ->
 - communauté disciplinaire
 - tutelles
 - financeurs

Contacts à PSL et à l'ENS-PSL

- Contacts institutionnels à l'ENS
 - Bibliothèques -> acces-ouvert@ens.fr / formations-bis-ens@ens.psl.eu / elif.su.sabuncu@ens.psl.eu
 - Service des archives -> lina.sbeih@ens.psl.eu
 - Délégué à la protection des données (DPD) de PSL -> donnees.personnelles@ens.psl.eu
(<https://psl.eu/politique-de-confidentialite>)
 - Service d'Accompagnement aux Appels à Projets (SAAP) -> saap@ens.psl.eu

Ressources utiles

1. Se former en ligne
2. Repérer les acteurs incontournables
3. Rédiger son DMP

1. Se former en ligne

- De très nombreuses formations...
- Doranum (<https://doranum.fr/>) : auto-formation, fiches,
- Institut biotechnologique de Flandres, Belgique <https://elearning.vib.be/courses/writing-a-data-management-plan/>
- Carnets de recherche (blogs) de chercheurs dont <https://ig.hypotheses.org/1450> (Guide pour plan de nommage et de classement) sur la description des documents figurés comme les photos.

2. Repérer les acteurs incontournables

- Le Comité pour la science ouverte : assure la mise en œuvre de la politique nationale de science ouverte (MESRI) -> <https://www.ouvrirlascience.fr/accueil/> (ressources +++)
- Le CNRS -> <https://www.science-ouverte.cnrs.fr/fr/> et DPO CNRS <https://www.cnrs.fr/fr/rgpd>
- L'ANR -> <https://anr.fr/fr/>
- Huma-Num -> <https://www.huma-num.fr/>
- La CNIL <https://www.cnil.fr/> : Textes, fiches pratiques, obligations des chercheurs concernant données personnelles
- La Section ESR de l'Association des archivistes français (AAF) -> <https://www.archivistes.org/sections/section-des-archivistes-des-universites-rectorats-organismes-de-recherche-et-mouvements-etudiants-aurore/>

3. Rédiger son DMP

- France : DMP Opidor : <https://opidor.fr/planifier/>
- Ailleurs : DMP online : <https://dmponline.dcc.ac.uk/>

à vous de jouer :

“La première chose que je vais mettre en
place **demain** est...”