



HAL
open science

Maîtriser les données de sa thèse : de la gestion à l'ouverture

Cécile Delay-Artous, Daphné Mathelier, Nathalie Reymonet, Aude Julien
Da cruz lima, Aurélie Gono

► To cite this version:

Cécile Delay-Artous, Daphné Mathelier, Nathalie Reymonet, Aude Julien Da cruz lima, Aurélie Gono. Maîtriser les données de sa thèse : de la gestion à l'ouverture. Doctorat. Maîtriser les données de sa thèse, Université Paris Nanterre, France. 2024. hal-04409017v1

HAL Id: hal-04409017

<https://hal.parisnanterre.fr/hal-04409017v1>

Submitted on 24 Jan 2024 (v1), last revised 28 Jun 2024 (v3)

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 Attribution 4.0 International License

Maîtriser les données de sa thèse

De la gestion à l'ouverture

Janvier 2024

Aude Da Cruz-Lima (UPN, DRED)

Cécile Delay (UPN, SCD)



Qui sommes-nous ?

- Aude Da Cruz-Lima
Chargée de valorisation, DRED – Université Paris Nanterre ;
adacruzlima@parisnanterre.fr
- Cécile Delay-Artous
Chargée de mission Données de la recherche, SCD – Université Paris Nanterre
c.delayartous@parisnanterre.fr
- ADN, l'atelier de la donnée de Nanterre, partenariat UPN, MSH Mondes et la Contemporaine
donnee-recherche@liste.parisnanterre.fr

Qui êtes-vous ?

Merci de renseigner ces informations dans le chat :

- Votre nom et prénom
- Votre école doctorale
- Votre discipline
- L'année de votre thèse
- Votre sujet de thèse



Programme

I. Introduction et définitions

- Science ouverte
- Science et société
- Données de la recherche
- Cycle de vie des données

II. FAIRisation et bonnes pratiques

- Principes FAIR
- Identifiants pérennes
- Métadonnées
- Bonnes pratiques

III. Cadre réglementaire

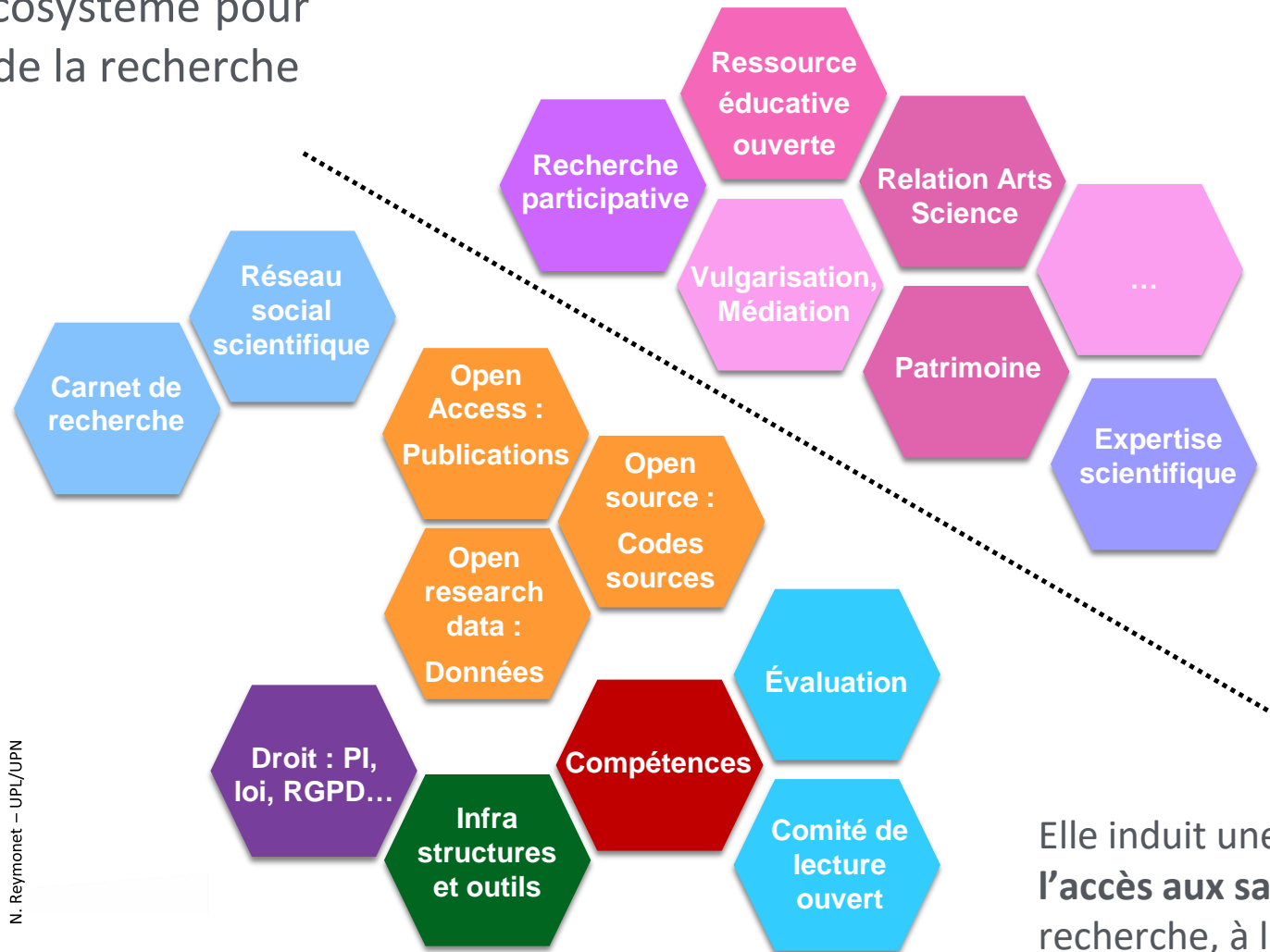
IV. Pour conclure : le PGD



I. Introduction et définitions

Introduction

La science ouverte : un écosystème pour un partage des résultats de la recherche



N. Reymonet – UPL/UPN

Elle induit une **démocratisation de l'accès aux savoirs**, utile à la recherche, à la formation, à l'économie, à la société.

Définition – 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.

Le mouvement de la Science ouverte vise à construire un écosystème dans lequel la science sera plus cumulative, plus fortement étayée par des données, plus transparente, plus rapide et d'accès universel.

Comité pour la science ouverte (2021). [Deuxième Plan national pour la science ouverte.](#)

Définition – Science et société

Actions de médiation, de communication ou de valorisation sur les enjeux et les résultats scientifiques en direction d'un public de non-spécialistes.

Elle se traduit par :

- Vulgarisation
- Médiation
- Expertise scientifique
- Patrimoine
- Relations entre arts et sciences
- Relations entre recherche publique et entreprises
- ...



Qu'est-ce qu'une donnée de recherche ?

Définition – Données de recherche

La définition la plus couramment utilisée est de 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](#) (2006)

Une définition plus large et plus opérante :

Il s'agit de l'ensemble des informations collectées, observées ou créées sous forme numérique dans le cadre d'un projet de recherche.

Deboin, M.-C. (2020). S'initier en ligne aux données de la recherche et à leur gestion. CIRAD.

<https://doi.org/10.18167/COOPIST/0055>

Diversité des données de la recherche



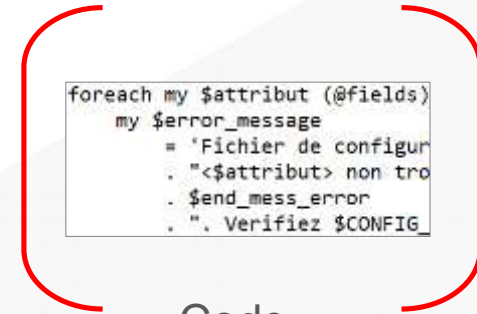
Textes



Archives



Manuscrits



Code



Enregistrements
sonores



Partitions



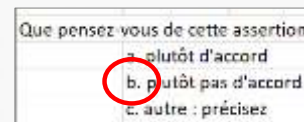
Photographies



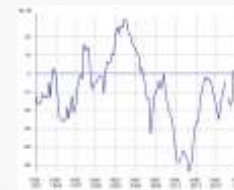
Cartes et plans



Spatiales ou
temporelles



Résultats
d'enquête



Séries
statistiques



Vidéos



Images 3D

etc.

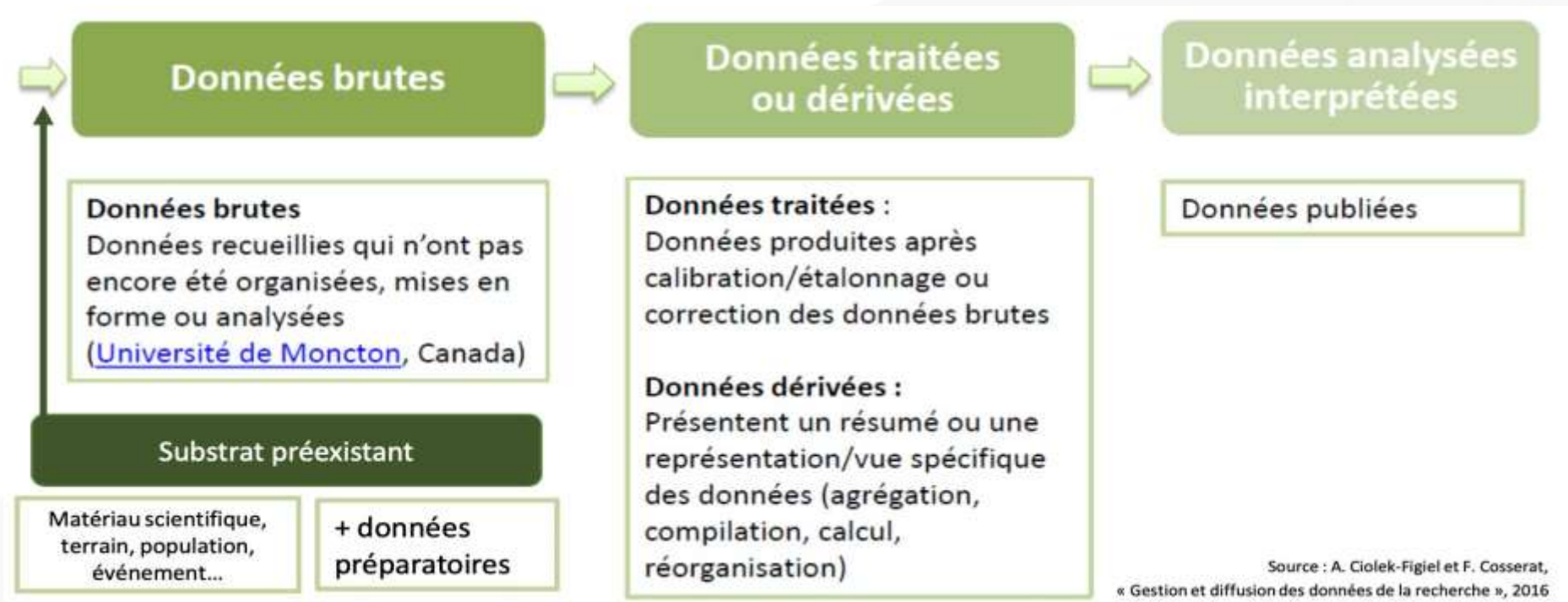
Données de recherche : une typologie

Types de données	Définitions	Valeur et spécificité	Exemples
Données d'observation (<i>Observational datas</i>)	Données obtenues en temps réel	Souvent uniques et irremplaçables	Données atmosphériques, d'enquêtes, échantillons, neuro-image
Données expérimentales (<i>Experimental datas</i>)	Données obtenues en laboratoire à partir d'équipements spécifiques	Reproductibles mais à des coûts dissuasifs	Séquence de génome, chromatographie, spectres RMN
Données de simulation (<i>Simulation datas</i>)	Données générées à partir de modèles test	Métadonnées et modèles ont une valeur supérieure aux résultats	Modèles climatiques, reconstitution en 3D de monuments historiques
Données dérivées ou compilées (<i>Derived or compiled data</i>)	Données obtenues par compilations ou traitement des données brutes	Reproductibles mais à des coûts dissuasifs	Texte et <i>data mining</i> , bases de données compilées
Données de référence ou données canoniques (<i>Reference or canonical datas</i>)	Collections statiques ou organiques de jeux de données validées	Données publiées ou qui ont fait l'objet d'une curation	Portail de l'Unesco sur les pays, banque de données sur le génome, structures chimiques

Donnée de recherche : des niveaux d'élaboration

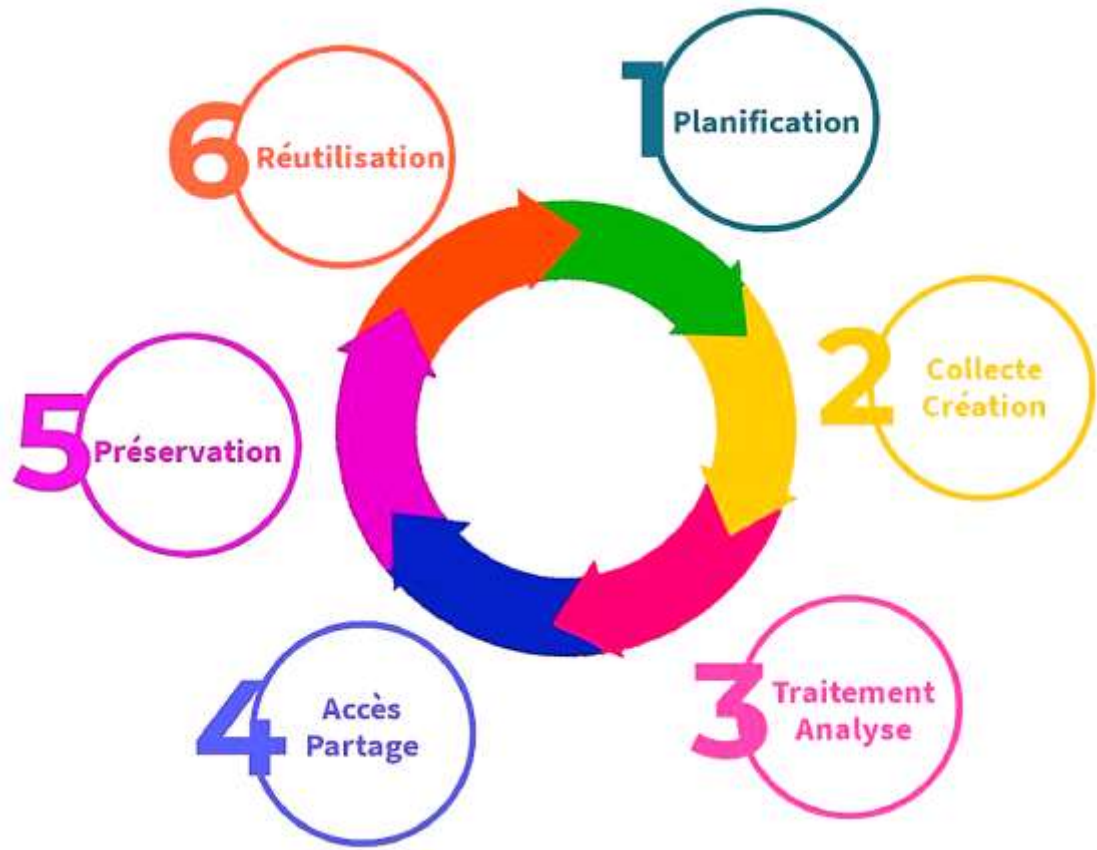
On peut aussi les définir en fonction de leur niveau d'élaboration :

- **Données brutes** (données collectées, non organisées, non mises en forme).
- **Données traitées** (organisées, structurées et prêtes à être analysées).
- **Données analysées** (telles qu'elles apparaissent dans les publications).



Cycle de vie des données scientifiques

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](#))



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

À chaque phase, des actions de gestion des données doivent être entreprises.

Gestion, partage, ouverture et archivage des données

Gérer

Mettre en place des **architectures**, des règles, des pratiques et des procédures pour satisfaire les besoins de la recherche tout au long du **cycle de vie** des données

Partager

Rendre accessible sur **autorisation** au moyen de protocoles maîtrisés, à des individus ou groupes identifiés, des données protégées par des exceptions légitimes encadrées par la loi : secret professionnel, secret industriel ou commercial, données personnelles ou contenus protégés par le droit d'auteur.

Ouvrir

Diffuser de manière structurée selon une méthode et une licence ouverte des données pour leur libre accès et leur réutilisation par tous, sans restriction technique, juridique ou financière

Archiver

Conserver à long terme des données qui présentent un intérêt scientifique, juridique ou historique



Pourquoi gérer ses données ?

Pourquoi gérer ses données ?

Pour ne pas risquer de perdre ou d'altérer ses données → **Stabilité des données**

Pour pouvoir les conserver à long terme → **Pérennité des données**

Pour améliorer la qualité et la finesse de ses données → **Qualité des données**

Pour garantir l'intégrité de la recherche → **Transparence et reproductibilité des données**

Pour pouvoir les partager plus facilement → **Réutilisation et visibilité des données**

Et (aussi)...pour aller dans le sens des orientations institutionnelles.

II. Gestion des données, FAIRisation et bonnes pratiques

FAIRisation des données

L'objectif des principes FAIR est de favoriser la **découverte**, l'**accès**, l'**interopérabilité** et la **réutilisation** des données partagées.

Chaque principe FAIR se décline en un ensemble de caractéristiques que doivent présenter les données et les métadonnées pour faciliter leur découverte et leur utilisation par les hommes mais aussi par les machines.

INRA (2018). [Produire des données FAIR.](#)

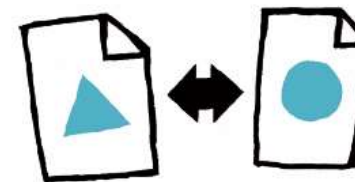
Ministère de l'Enseignement supérieur et de la Recherche (2022).

[Passeport pour la science ouverte](#)



Le principe **Findable** (Facile à trouver) a pour objectif de faciliter la découverte des données par les humains et les systèmes informatiques et requiert une description et une indexation des données et des métadonnées.

Le principe **Accessible** encourage à stocker durablement les données et les métadonnées et à faciliter leur accès et/ou leur téléchargement, en spécifiant les conditions d'accès (accès ouvert ou restreint) et d'utilisation (licence).



Le principe **Interopérable** peut se décomposer en : téléchargeable, utilisable, intelligible, et combinable avec d'autres données, par des humains et des machines.

Le principe **Reusable** (Réutilisable) met en avant les caractéristiques qui rendent les données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction/transparence de la science).



Identifiants pérennes

Numéro ou une entité alphanumérique, permettant de désigner et de **retrouver de manière univoque et pérenne un objet, un document, une personne, un lieu, un organisme**, ou toute entité, dans le monde physique ou numérique.

Des identifiants ouverts pour la science ouverte : note d'orientation du [Comité pour la science ouverte](#) (2019)

Un identifiant pérenne ou PID (persistent identifier) est :

- Unique : il se réfère à une seule ressource
- Pérenne : sa durée est garantie dans le temps
- Résolvable : il permet d'accéder à la ressource ou au moins à une description de celle-ci

Identifiants pérennes objet

Plusieurs identifiants objet existent mais c'est le DOI qui est recommandé.



10.5072 / datacenter.123xy

Préfixe Suffixe

DOI = Digital Object Identifier

Certains entrepôts de données attribuent automatiquement un identifiant pérenne aux données, par exemple :



ID : 10.34847/nkl.7dcdw459



DOI 10.5281/zenodo.4291876

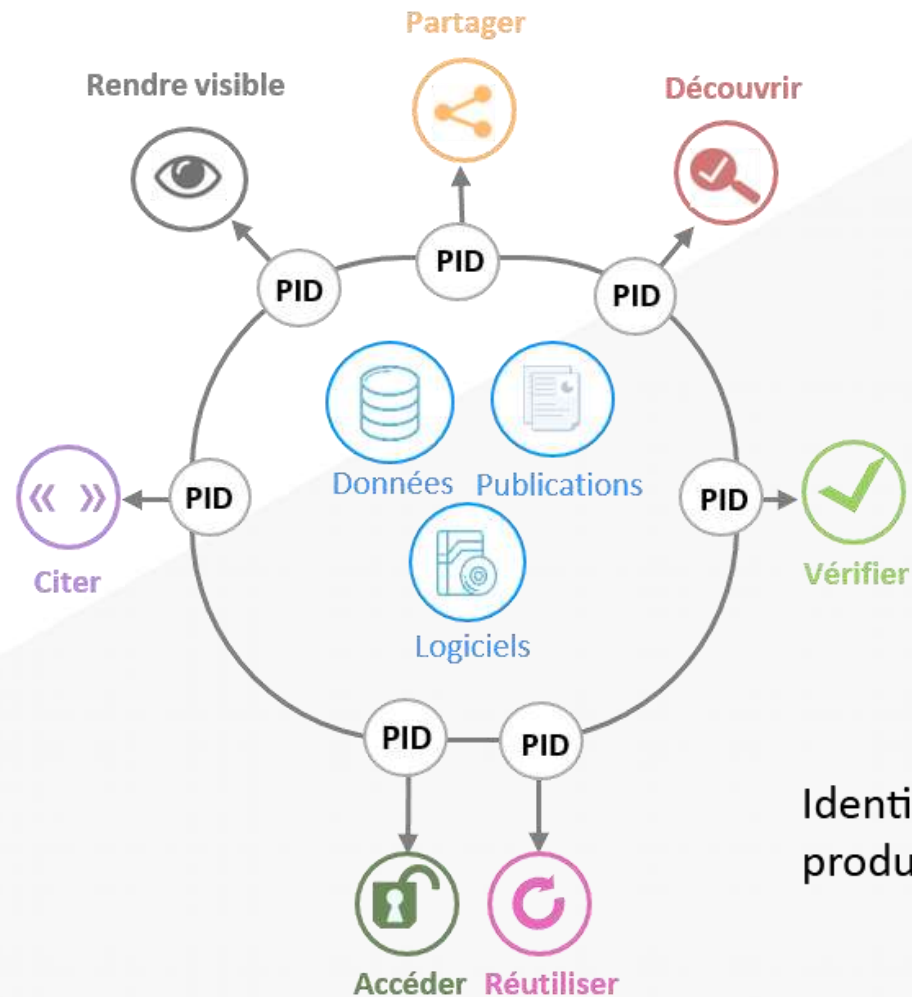


Identifiant pérenne du dataset ?

doi:10.57745/NTSGDF

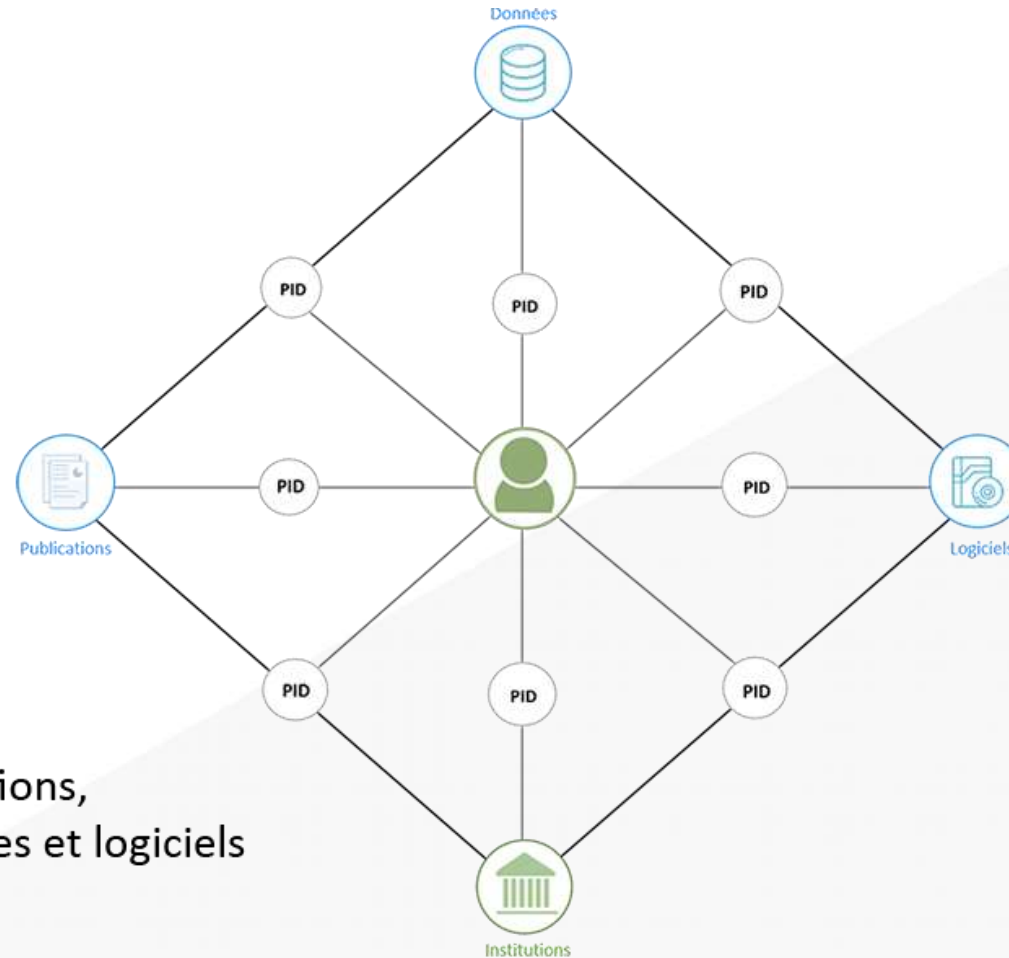
Identifiants pérennes objet

Le DOI est l'identifiant de ressources les plus utilisés : plus de 130 millions de DOI en 2017



Identifier et retrouver les productions scientifiques

Identifiants pérennes contributeur



Lier auteurs, institutions,
publications, données et logiciels

ORCID est
l'identifiant le plus
utilisé dans le monde
avec plus de 10
millions de comptes
en 2019. On peut le
lier à d'autres
identifiants auteur.

DoRANum (2020). [Les identifiants pérennes](#)
DOI : 10.13143/t427-f432

Métadonnées



Ensemble structuré d'informations décrivant une ressource quelconque, numérique ou non.

[Enssib](#)

=> C'est **une donnée sur une donnée.**



Métadonnées – Pourquoi faire ?

- Comprendre l'origine des données et leur contexte de création ou de collecte
- Améliorer le moissonnage par les machines (moteur de recherche)
- Garantir l'interopérabilité
- Connaître les conditions de réutilisation et de partage des données
- Fournir des informations très utiles quand les données ne peuvent pas être partagées (embargo, accès restreint) ou lors du retrait de données (données obsolètes, etc.)

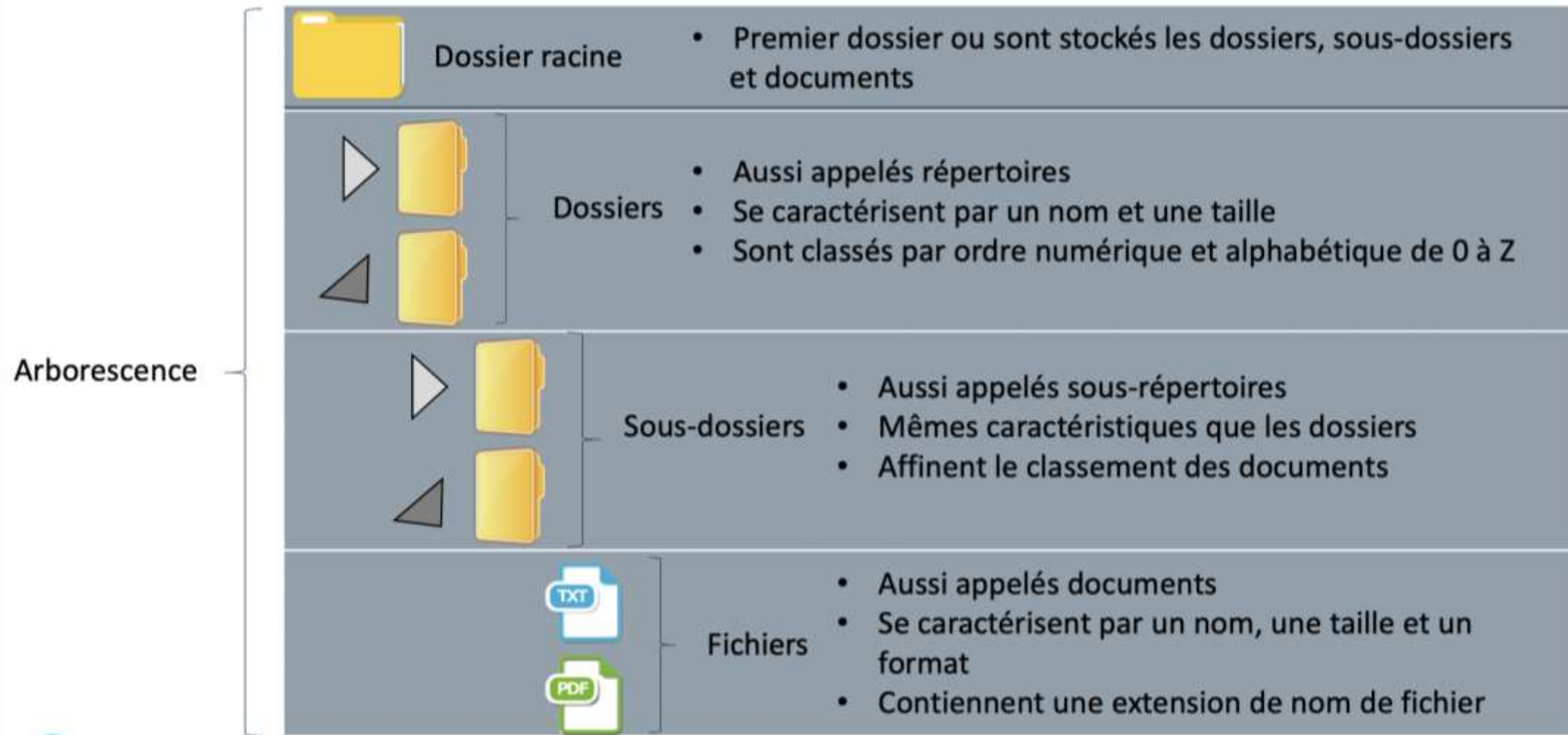
Bonnes pratiques de gestion des données

- Réfléchir à l'**organisation** de ses fichiers de données
- **Décrire** et **documenter** ses données de façon à les identifier sans erreur (nommage, versionnage...)
- Choisir des **formats** de fichier adaptés pour la conservation à long terme
- Opter pour une solution de **stockage** adaptée
- Planifier la **sauvegarde** de ses données
- Anticiper un éventuel **partage des données**
- Réfléchir aux **aspects éthiques et juridiques**

Organiser ses fichiers

- Il n'y a pas vraiment de « bonne » **arborescence**, l'essentiel c'est de s'y retrouver et que le classement fasse sens.
- On peut organiser les documents par projet, action, date, type de données...
- Numéroté éventuellement vos dossiers
- Séparer les données brutes des données traitées ou analysées.
- L'arborescence des dossiers ne doit pas dépasser 5/6 niveaux.
- Appliquer la même logique de classement dans tous vos espaces de stockage.
- Si votre système de classement ou vos données complexes, documenter vos choix dans un document (du type *Readme.txt*)

Arborescence



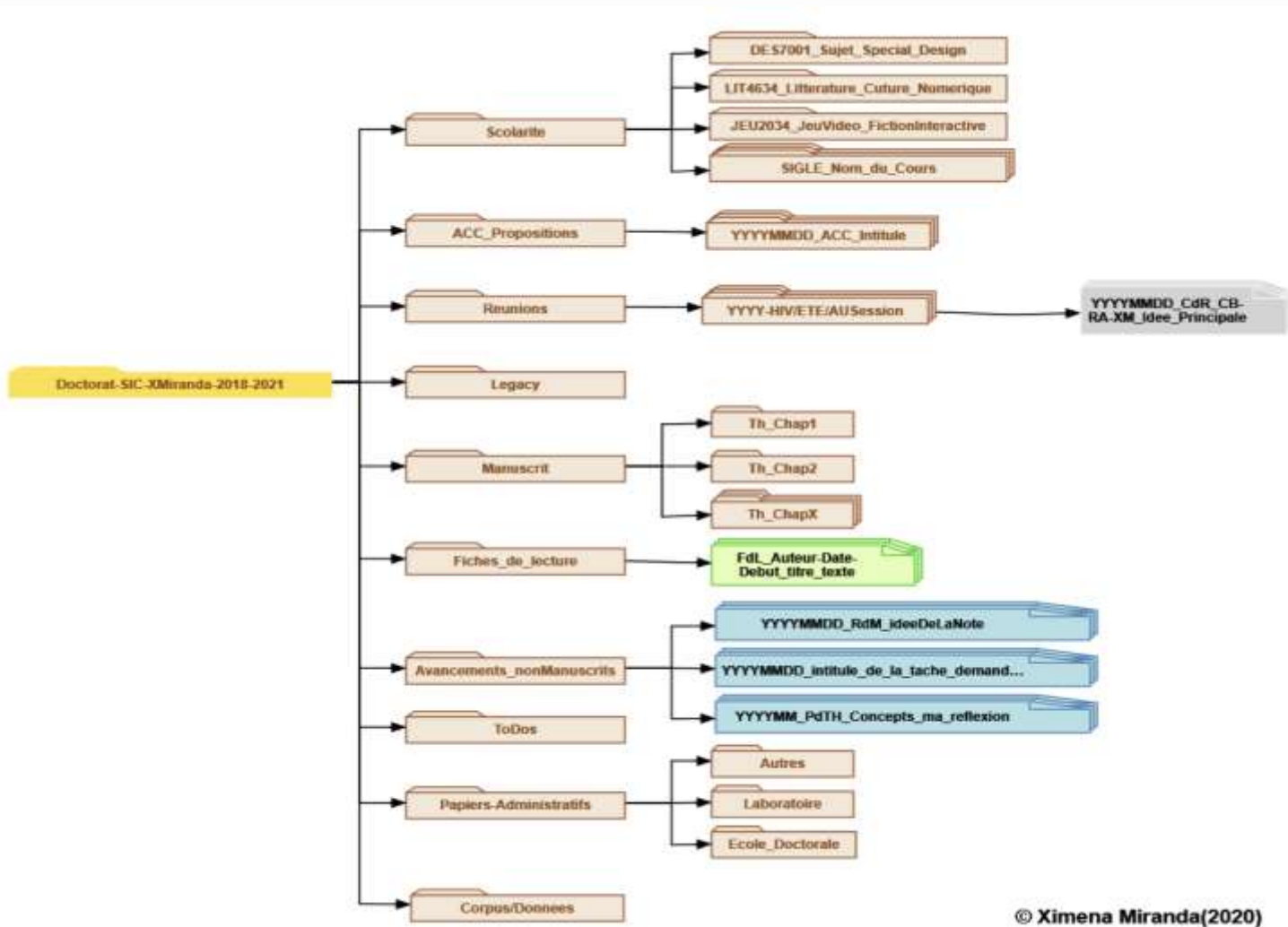
Nommer ses fichiers

Cinq grands principes ([DoRANum](#)) :

1. Donner un nom bref et explicite
2. Ne pas introduire d'espace ou de caractère accentué
3. Formater les dates et les nombres
4. Définir clairement les éléments importants et leur ordre d'apparition
5. Préciser éventuellement les versions du doc



Un exemple d'organisation



Dossier type de doctorat proposé par Ximena Miranda (2020) à partir du logiciel [ThinkComposer](#).

Voir le billet dédié à cette question sur son [blog](#).

Documenter ses données

Dans la mesure du possible, expliquer comment les données ont été obtenues, transformées, comment elles sont organisées :

- À l'aide de fichiers textes lisibles par des humains (ex. fichier *Readme.txt*)
 - Trois niveaux de description : projet, jeux de données et données
- En renseignant les métadonnées
 - Au fur et à mesure de l'[avancée du projet](#), en particulier au moment de partage des données et lors de l'archivage pérenne.
 - En utilisant un [standard de métadonnées](#) propre à votre discipline (ex. DDI) ou reconnu par la communauté universitaire (ex. Dublin Core).
- En rédigeant un [data paper](#) pour expliquer et donner accès aux données

Choisir des formats adaptés

- Format informatique = convention pour représenter une donnée sous forme numérique.
- Différents types de formats : format ouvert, format fermé, format normalisé, format standard, format propriétaire, format conteneur...
- Format « idéal » : ouvert, normalisé, largement diffusé, compatible avec de nombreux logiciels, facile à transformer et indépendant.








BU d'Aix Marseille Université, DoRANum (2018). [Liste indicative formats ouverts](#)

Urfist Méditerranée, DoRANum (2017) Jeu sur les formats : [Format ouvert ou fermé ?](#)

DOI : 10.13143/mcwq-qs64

Choisir des formats ouverts

Il est important de privilégier les formats ouverts pour faciliter le partage et l'archivage des données.

	Archivage	Image	Dessin	Audio	Texte	Tableur	Vidéo
Formats fermés							
Formats ouverts							

DoRANum, Urfist Méditerranée (2017) Jeu sur les formats : [Format ouvert ou fermé ?](#)

DOI : 10.13143/mcwq-qs64

Choisir une solution de stockage

Estimer la volumétrie des données et choisir une solution d'hébergement :

- Préférer idéalement un espace de stockage institutionnel à une solution commerciale (pour garder la main sur ses données).
- Éviter d'avoir recours aux périphériques externes (disques durs, clés USB) car ils sont fragiles et faciles à perdre.
- Privilégier les stockages permettant la synchronisation des données.

DoRANum (2019), [Stockage, partage et archivage : quelles différences ?](#)

DOI : 10.13143/5dax-qp58

Grille de lecture des solutions de stockage

Support de stockage	Sécurité	Accès	Coût	Remarque d'utilisation
 Ordinateur professionnel	★★☆☆ Sujet au piratage informatique, aux détériorations et pannes.	★☆☆☆ Pas adapté au partage, nécessite l'utilisation d'un support externe ou d'Internet (mail, cloud...)	★★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter les données confidentielles et sensibles
 Support externe	★☆☆☆ - Sujet au vol, à la perte du support - Durée de vie limitée (dégradation du matériel)	★★★★★ Facilement transportable, il permet de transférer les données vers un autre ordinateur	★★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter ou de sécuriser physiquement les données confidentielles et sensibles
 Serveur institutionnel	★★★★★ Stockage fiable, durable et sécurisé (contre le vol, le piratage, les incendies...)	★★☆☆ La connexion au serveur institutionnel ne facilite pas le travail avec des personnes extérieures	★★☆☆ Coût assez important mais pas forcément répercuté sur l'utilisateur	- Pour un stockage plus pérenne - Adapté pour le stockage de données sensibles et des versions « stables » de vos données - Toutes les institutions ne proposent pas ce service
 Serveur Cloud	★★☆☆ On ne sait pas vraiment où sont stockées les données, ni ce qu'elles deviennent	★★★★★ Permet un travail synchronisé avec toutes les personnes ayant été autorisées au partage	★★☆☆ Payant à partir d'une certaine limite de stockage	- Pour un partage avec des personnes externes à l'institution - Ne pas y mettre de données sensibles ou confidentielles - Pas de contrôle sur la procédure de sauvegarde des données



Et vous, comment stockez-vous vos données ?

Quelles solutions de stockage à Nanterre ?

Espaces ouverts aux doctorants :

- [Huma-Num](#) : **ShareDocs** (90 Go) => accès en créant un compte HumanId ou via eduGain
 - [CloudMe](#) (3 Go) => accès en créant un compte CloudMe (option recommandée pour Zotero)
 - La [DRI](#) de l'Université Paris Nanterre donne accès plusieurs suites collaboratives :
 - [Office 365](#) : **OneDrive**
 - [Google Workspace for Education](#) : **Google Drive**
 - Open 365 (en cours de déploiement)
- => Accès avec vos identifiants UPN

Choisir une solution de sauvegarde

- Sauvegarde efficace = duplication et stockage des données à différents endroits, sur différents supports, à différents moments.
- [Règle du 3-2-1](#) : garder **3** exemplaires des données, sur **2** supports ou technologies différents, dont **1** se trouve hors site.
- Réfléchir à une stratégie de sauvegarde, notamment lors des points d'étape d'un projet, et gérer les différentes versions (sans trop les multiplier).

Trois phases de sauvegardes




Choisir une solution de partage

Entrepôts de données = service en ligne permettant la collecte, la description, la conservation, la recherche et la diffusion des jeux de données. [IRDData](#), Entrepôts de données.

- Entrepôts disciplinaires : ex. [Nakala](#) pour les SHS
- Entrepôts institutionnels : ex. instance UPN de [RechercheDataGouv](#)
- Entrepôts généralistes : ex. [Zenodo](#)
- Entrepôts liés : ex. [Cybergeos Dataverse](#) (revue Cybergeos)
- Annuaire d'entrepôts : ex. [Re3Data](#)

Réutiliser des données

1. Repérer un jeu de données



- Dans une publication 
- Dans un moteur de recherche, une BDD, un entrepôt, etc.



2. Réutiliser ces données dans sa recherche

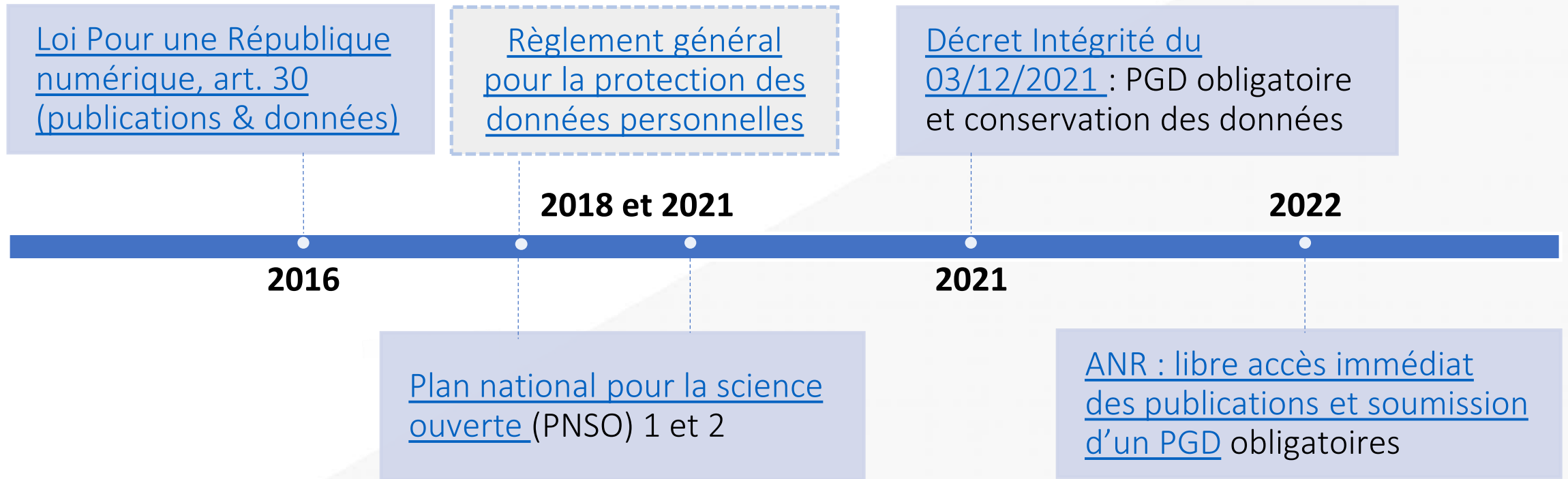
- Vérifier les conditions de réutilisation 

3. Citer correctement ce jeu de données

- Mentionner l'auteur, son identifiant 
- Mentionner l'identifiant du jeu de données 

III. Cadre réglementaire

Évolution de la réglementation





Selon vous,
à qui appartiennent vos données ?

Propriété des données

Dans leur activité de recherche académique, les chercheurs (enseignants-chercheurs, chercheurs, doctorants, autres personnels de la recherche) produisent des données « publiques » :

- Ils/elles n'ont pas de propriété intellectuelle sur les données factuelles collectées (uniquement sur les œuvres de l'esprit, le critère étant l'originalité).
- Les établissements « producteurs » ont un droit de propriété intellectuelle sur la structuration des données qui donne lieu à une base de données, mais pour les établissements publics, c'est l'ouverture par défaut qui prévaut (Loi pour une République numérique, art. 11).

La question de la propriété des données n'est pas forcément le bon angle d'attaque depuis la Loi pour une République numérique (et le principe d'ouverture par défaut). Aujourd'hui, savoir qui est le propriétaire des données n'est souvent plus nécessaire. Il s'agit de savoir si on est dans le principe de l'ouverture par défaut ou bien dans un cas d'exceptions.

Licences ouvertes

Une licence ouverte garantit à tous le droit d'utiliser, de partager et d'accéder à vos données avec la sécurité juridique nécessaire aux producteurs et aux ré-utilisateurs des données.














DoRANum (2022), [Guide des licences ouvertes](#) DOI : 10.13143/tv6f-sv31

Principales licences ouvertes :

- Licences Creative Commons
- Licence Etalab
- Licences logicielles (GNU, BSD, Apache, MIT, CeCILL)
- Licences bases de données (ODbL, PDDL, OCD-By)

CopIST, Dedieu, L. (2015), [Rendre publics ses jeux de données scientifiques](#)

Les licences ouvertes Creative Commons

		Utilisation Partage	Adaptation Modification	Utilisation commerciale	Modification de licence	
TRÈS LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
LIBRE						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
NON LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix NON permis
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix NON permis



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

A quoi ressemblent les CC ?

Replication Package for Immigration and Redistribution

Alessia, Alberto; Milano, Armando; Stantcheva, Stefania

Replication package for Immigration and Redistribution by Alberto Alessia, Armando Milano, and Stefania Stantcheva.

The package contains all the survey data and the data on actual statistics about immigrants and non-immigrants, as well as the scripts needed to replicate the figures and tables in the paper. The package also contains three excel files detailing how we computed the actual statistics and the sources we used.

Files listed in the preview:

Do	8.2 kB
DS_Store	
Figures	
Fig10.do	11.0 kB
Fig2to7.Rmd	40.3 kB
Fig2to7_0_data.do	48.1 kB
Fig8.do	8.1 kB
Fig9.Rmd	4.3 kB

zenodo

Indexed in OpenAIRE

Publication date: January 18, 2022

DOI: [10.5201/zenodo.5838806](https://doi.org/10.5201/zenodo.5838806)

License (for files): Creative Commons Attribution 4.0 International

License (for files):

 Creative Commons Attribution 4.0 International

HAL SHS sciences humaines et sociales

Accueil | Déposer | Parcourir | Rechercher | Services | Documentation HAL

hal-0271307, version 1

Carpooling and carsharing for commuting in the Paris region: A comprehensive exploration of the individual and contextual correlates of their uses

Jules Bellon, Thierry Perle, Sigrid Denon

UNIQ - Université de Versailles Saint-Quentin-en-Yvelines
CEPR - Culture, Environnement, Aménagement, Représentations, Climat
LPA - Université Paris 8 Vincennes-Saint-Denis
LADYSS - Laboratoire Dynamiques Sociales et Innovations des Espaces

Type de document: [Communication](#)

Domaine: [Sciences de l'Économie et du Social](#) - [Économies et Finances](#) - [Sciences de l'Économie et du Social - Économie](#) - [Sciences de l'Économie et du Social - Méthodes et Modèles](#)

hal-0271307, version 1

HAL SHS

CC BY

LICENCE



Distributed under a Creative Commons Paternité - Pas d'utilisation commerciale 4.0 International License

Comment savoir quand ouvrir ?

Depuis la *Loi pour une République numérique*, **ouverture par défaut** des données de la recherche, sauf exceptions (*nombreuses, voir les vignettes suivantes*).

Ces données doivent être diffusées en **libre accès** (ou a minima être **communicables à la demande**) et ce, **gratuitement**.

Lionel Maurel (2021), [Quel cadre pour l'ouverture des données de la recherche dans le contexte de la science ouverte ?](#)

En cas de doute, il existe plusieurs outils d'aide à la décision :

- [Arbre de décision](#) du CIRAD
- [Logigramme](#) de l'École des Ponts ParisTech
- [Logigramme](#) de l'INRAE
- [Logigramme](#) de l'Institut Pasteur

et pour savoir où déposer ses données : un [logigramme de Recherche Data Gouv](#).

Les exceptions à l'ouverture : les données à caractère personnel

Règlement général sur la protection des données (RGPD) : protection des personnes physiques à l'égard du traitement des données personnelles et à la libre circulation de ces données.

= toute information relative à une personne physique susceptible d'être identifiée, directement ou indirectement.

→ **Anonymisation** : traitement qui vise à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière irréversible (pas de réidentification des personnes par recoupement).

→ **Pseudonymisation** : traitement qui rend impossible l'attribution des données relatives à une personne physique sans information supplémentaire : remplacer les données directement identifiantes (nom, prénom, etc.) d'un jeu de données par des données indirectement identifiantes (alias, numéro séquentiel, etc.).

Contact : dpo@liste.parisnanterre.fr

Les autres exceptions à l'ouverture

- Secret **médical**
- Secret **industriel** ou commercial : influencer la position sur un marché ou révéler des procédés de fabrication
- Secret statistique
- Sécurité des systèmes d'information des **administrations**
- Secret **défense**, sûreté de l'État (données stratégiques), sécurité publique
- Données sur lesquelles des tiers ont **priorité d'exploitation**
- Données sur lesquelles des tiers ont des droits de **propriété intellectuelle**

Ouverture raisonnée des données

Le principe « Aussi ouvert que possible, aussi fermé que nécessaire »
reste la règle dans le domaine de la recherche.



Qu'est-ce qu'un plan de gestion de données
et qui cela concerne-t-il ?

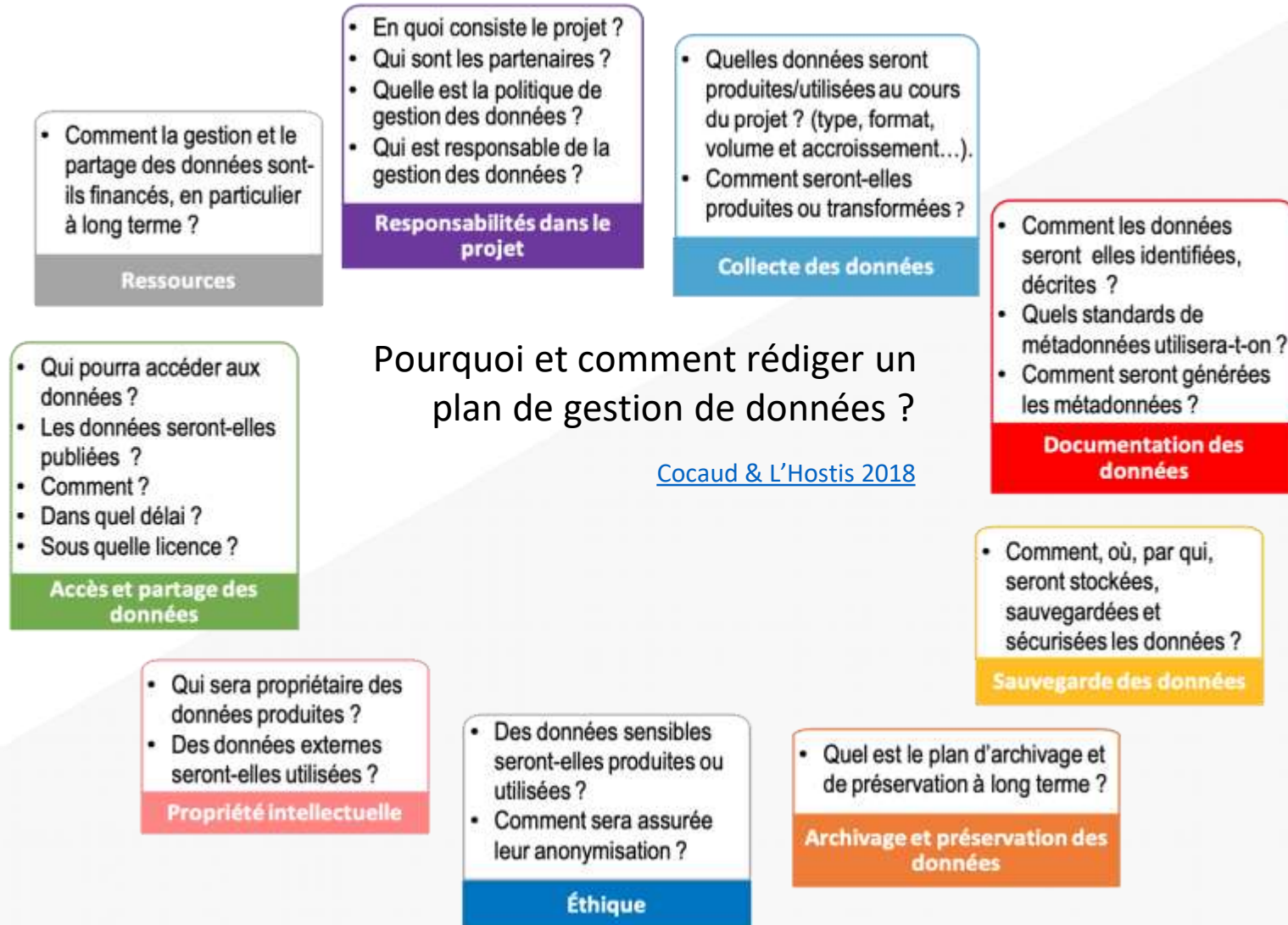
Le plan de gestion de données

Le plan de gestion de données (PGD) est un document formel et évolutif précisant la manière dont les données seront produites, traitées, décrites, partagées ou protégées et conservées au cours et à l'issue du projet.

Objectifs

- Décrire le cycle de vie des données produites ou collectées au cours du projet de recherche.
- Anticiper les questions de gestion qui surviennent au cours d'une recherche et les conditions d'une diffusion et d'une conservation futures des données.

Les grandes rubriques d'un PGD



Ressources

- CIRAD (2016), [Le cycle de vie des données. Intégrer la gestion de données scientifiques aux activités de recherche](#) (poster)
- Deboin M.-C. (2020). [S'initier en ligne aux données de la recherche et à leur gestion](#) (guide du CIRAD)
- [DoRANum](#), modules de formation, quiz et tutoriels sur la gestion et le partage des données de recherche, notamment [Le parcours interactif sur la gestion des données de la recherche](#)
- Huma-Num (2021), [Gérer et diffuser ses données avec les outils et services proposés par Huma-Num](#) (supports de formation et vidéos)
- Lionel Maurel (2021), [Quel cadre pour l'ouverture des données de la recherche dans le contexte de la science ouverte ?](#) (séminaire Datasuds 2021)
- Ouvrir la science (2020), [Passeport pour la science ouverte. Guide pratiques à l'usage des doctorantes et doctorants](#) (guide)
- Urfist de Paris (2021), [Introduction aux données de la recherche](#) (support de formation)
- [DMP OPIDoR](#), plateforme d'accompagnement à l'élaboration et la mise en pratique de plans de gestion de données et de logiciels.

MERCI !

Support réalisé par

Nathalie Reymonet

Cécile Delay

Daphné Mathelier

Aude Da Cruz Lima

Université Paris Nanterre / MSH Mondes/ADN

