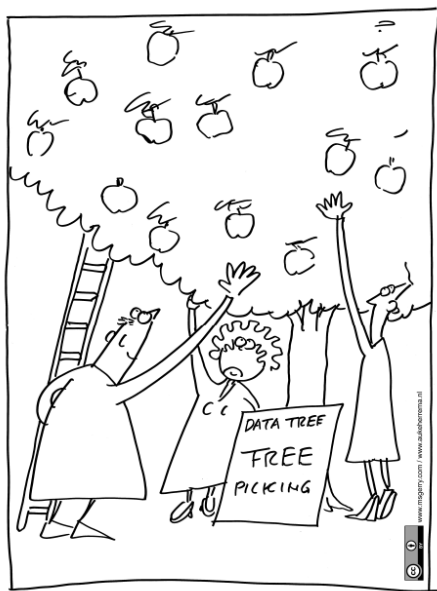


Introduction générale aux données de la recherche

BU
Lisa Malguy
Drive
Gemma Davies

09/12/2022



THE FUTURE?

Journées Science ouverte
8 et 9 décembre 2022

Définir ce qu'est une donnée

- Variété de formes et de perceptions en fonction du champ disciplinaire
 - On peut la définir selon
 - sa nature (image, son, texte, etc.);
 - son origine (collectée/réutilisée);
 - son degré de traitement (brute, nettoyée, analysée ou primaire/secondaire, chaude/froide);
 - son mode d'obtention (donnée d'observation, donnée de simulation, donnée expérimentale);
 - sa qualification (quantitative/qualitative).
 - Ou la définir par ce qu'elle n'est pas (publication, carnet de laboratoire, donnée administrative)...

Définir ce qu'est une donnée

- C'est le chercheur qui érige une entité en donnée par un arbitrage scientifique
- Donnée = valeur de preuve scientifique
- en SHS on parlera plutôt de la production qui sert à étayer une argumentation

Définir ce qu'est une donnée

La définition qui fait consensus est celle de l'OCDE en 2007 :

« [...] 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. [...] Ces Principes et Lignes directrices portent essentiellement sur les données de la recherche sur support numérique exploitable sur ordinateur. »

« *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics* ». Disponible sur : <http://www.oecd.org/fr/science/inno/38500823.pdf>

Enjeux?

- Importance du contexte technique
 - Ère des données massives (Big data) et de la fouille de texte en lien avec l'intelligence artificielle
 - Développement des méthodes de découvertes scientifiques par des corrélations sur de grandes masses de données
- Tendances = valeur données décorrélée de la publication

Enjeux?

- Intégrité scientifique
 - Contexte concurrence internationale avec multiplication des fraudes scientifiques qui aboutit à un besoin de transparence
 - Accès aux données dans un souci de reproductibilité de la science
 - Renforcer la confiance des citoyens dans la science
- [Référent Intégrité scientifique de l'UBO](#)

Enjeux?

- Rentabiliser les investissements publics
 - Eviter les pertes de données
- Favoriser les collaborations
 - Eviter les redites : réutiliser les données existantes plutôt que les recréer
 - Enjeux économiques : transferts des connaissances pour innover
 - Favoriser l'interdisciplinarité
 - Difficulté à trouver un langage commun dans un contexte de collaboration
 - Potentiel de réutilisation?

Cadre

- Engagement politique français
 - Deuxième axe du [Plan national pour la science ouverte de 2021-2024](#)
 - Mise en place d'infrastructures, comme [Recherche Data Gouv](#) en juillet 2022, et de services d'accompagnement dans la perspective de rendre obligatoire l'ouverture des données de la recherche dans la limite des exceptions à la Loi
 - S'appuie sur la [Loi pour une République Numérique de 2016](#)
 - Établit un principe d'ouverture par défaut des données publiques. Or les données de la recherche sont considérées comme des données publiques. Ce sont les établissements de recherche qui sont propriétaires des données et non le chercheur

Cadre

- [Décret n°2021-1572](#) : Obligations des établissements sur la gestion des données de recherche
 - conservation des données réalisés en son sein (pour la vérification)
 - Définition d'une politique de conservation, communication et de réutilisation de ces données
 - Veiller à la mise en œuvre de PGD par son personnel

Cadre

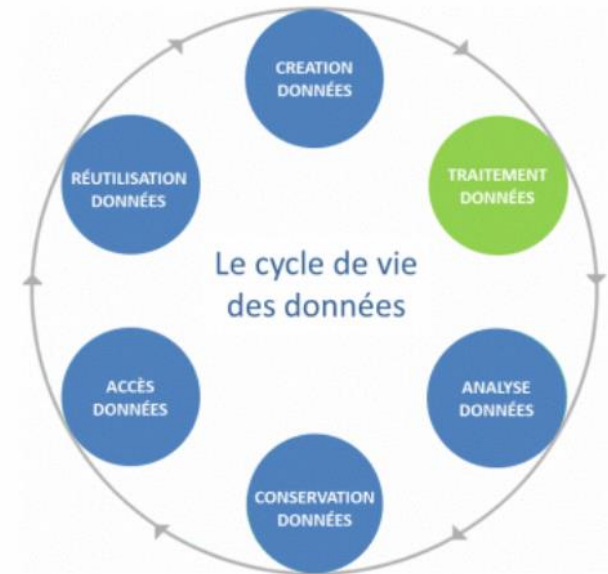
- Rôle des financeurs publics et des revues
 - Le financement de la recherche sur projet par Horizon Europe ou l'[ANR](#) impose la livraison d'un Plan de Gestion de Données (PGD/DMP) depuis 2019
 - + Incitation à le rendre accessible dans un entrepôt (ex. Zenodo)
 - De plus en plus de revues recommandent/exigent l'accès aux données pour la publication d'un article (*supplementary materials*)
 - Exemple [PLOS](#), [Springer Nature](#)

Obligations de partage?

- Ce qu'il faut retenir :
 - Le partage des données n'est pas obligatoire mais fortement recommandé lorsque cela est possible
 - « **Aussi ouvert que possible, aussi fermé que nécessaire** »
 - En revanche, ce qui est obligatoire c'est de remplir un PGD
 - Dans ce PGD, il faudra justifier du pourquoi on ne partage pas ses données

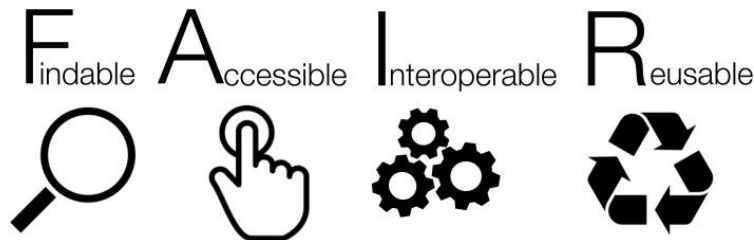
Le cycle de vie des données : distinguer gestion/conservation/partage

- Gestion et sécurisation des données, un impensé
- Rendre les données accessibles sur le long terme est indépendant de les rendre publiques
- Vers une **partage sous contrôle** : avant tout, rendre fiable le circuit des données



Modèle UK Data Archive

Répondre aux principes

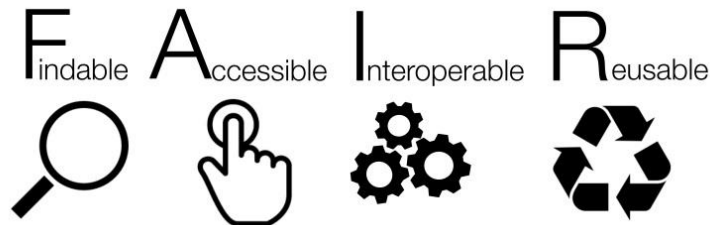


- Recourir à des **standards** pour structurer ses données et ses **métadonnées**
 - Utiliser des formats de fichiers ouverts et des règles de nommage des fichiers et de versioning
 - Utiliser un vocabulaire « contrôlé » pour exprimer des métadonnées riches
 - Associer des identifiants pérennes aux jeux de données (DOI), aux auteurs (ORCID), aux établissements (RNSR/ROR), etc.
 - *A minima*, indiquer provenance des données et une licence d'utilisation (CC0)
 - Mais aussi : le contexte de leur production, les outils nécessaires

Quelles licences pour les données?

- Licences applicables
 - Datalab : 2011, par l'administration française.
 - ODBL : 2010, pour les bases de données
 - CC0 : ajouter une phrase pour préciser qu'il faut citer la source
- Les licences creative commons ne sont pas adaptées aux données

Répondre aux principes



- Choisir un **entrepôt de données ouvert**
 - Stockage ≠ [archivage](#)
 - Gouvernance? Public (Nakala) /Privé (Figshare)? Enjeux de souveraineté
 - Généraliste/institutionnel/thématique
 - National (Nakala) /International (Zenodo)
 - Caractéristiques?
 - Certification? [CoreTrustSeal](#) : organisme de certification qui garantit aux déposants que leurs données seront protégées et gérées de manière optimale

→ Répertoire pour sélectionner son entrepôt : [Re3data](#)

Choisir un entrepôt

En fonction objectifs d'impact du projet

Respect des principes FAIR

Critères de choix

Exigences de l'organisation, du financeur, de l'éditeur

Communauté d'appartenance

Réputation de l'entrepôt

Accès contrôlé Embargo?

Localisation du serveur

Facilité de dépôt

Garanties de pérennité

Volumétrie acceptée

Coût?

Licences proposées

Formats acceptés Interopérabilité

Richesse métadonnées

Attribution identifiants type DOI

type de données acceptées

Imposées par le chercheur/institution

Imposées par l'entrepôt

Voir cours sur [Callisto](#)

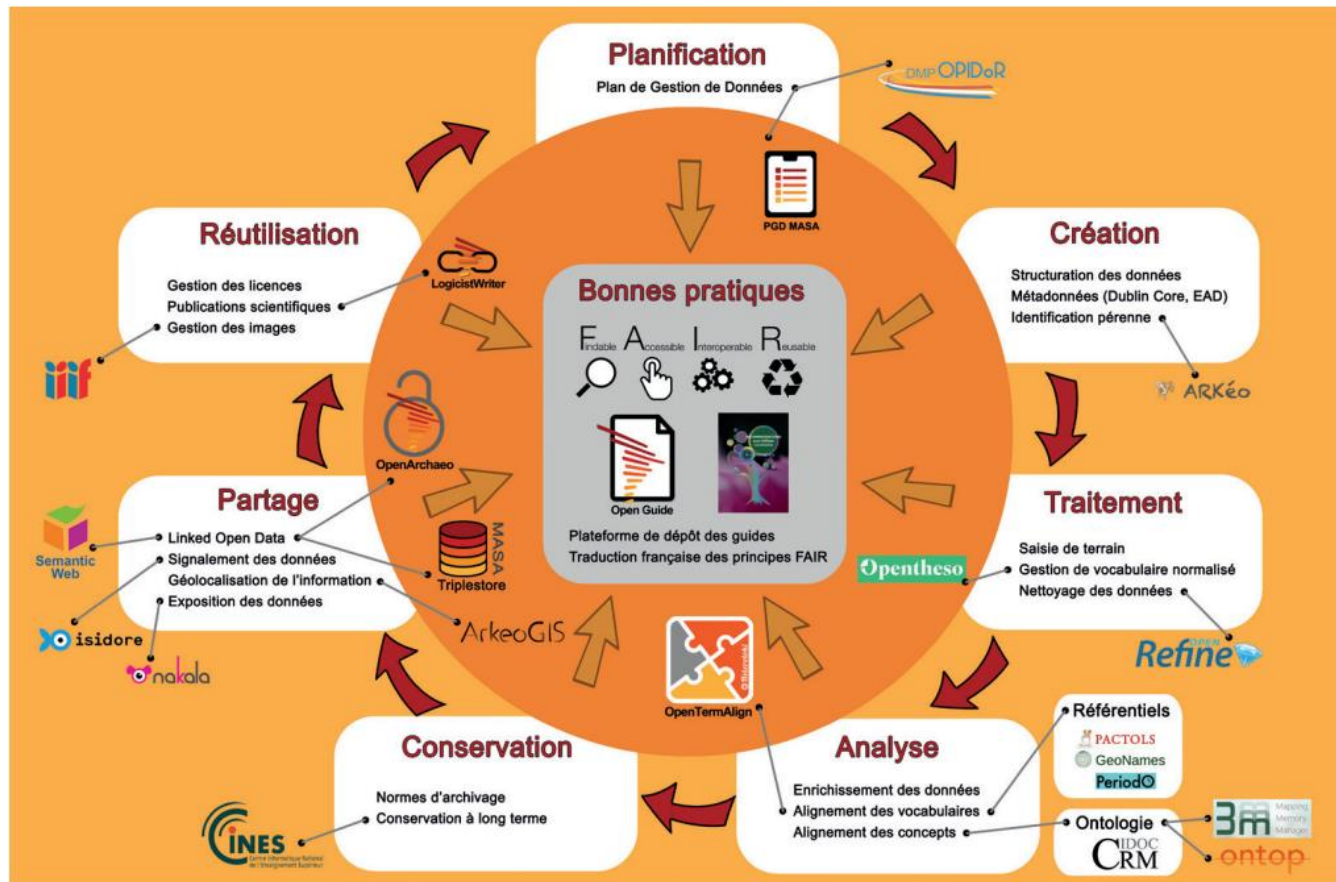
Extrait poster du Cirad

Un autre exemple de gestion organisé à partir du cycle de vie des données : à chaque étape des services spécifiques



Huma-Num Archéologie

Un exemple de gestion organisé à partir du cycle de vie des données : à chaque étape des outils et des ressources spécifiques



Ecosystème numérique MASA

Rédiger un plan de gestion de données

- Le PGD est un document évolutif qui suit le cycle de vie des données (doit donc être réactualisé au fil du projet)
 - Il faut le commencer le plus en amont possible pour **anticiper** les problématiques
- Synthétise la façon dont vous allez gérer vos données tout au long de votre projet et après
 - Formalisation au sein d'un document unique d'éléments dispersés entre divers acteurs et divers documents
- Organisation sous forme de questions
 - Il oblige à se poser des questions méthodologiques, juridiques et techniques

Avantages de la rédaction d'un PGD

- Outil de réflexion et de planification
 - D'abord un outil de gestion interne qui permet de :
 - se ménager la possibilité de ré-exploiter ses données
 - de faciliter la communication en tout premier lieu avec ses collaborateurs, ses encadrants...
- Facilitateur à long terme
 - Pour attester les résultats obtenus devant les pairs
- Accroît l'impact de votre recherche. Vos données gardent une valeur à long terme
 - A noter : Si vous changez d'établissement, l'accès à vos données est garanti

→ Aller plus loin [en rédigeant un Data Paper!](#)

Pour intégrer dans le processus de recherche l'action de diffusion des données comme un produit de recherche à part entière, évalué par les pairs et citable

DMP OPIDoR



Data **M**anagement **P**lan **O**ptimisation du **P**artage et de
Interopérabilité des **D**onnées de **R**cherche

Accessible : à l'ensemble de la communauté et à ses partenaires
français ou étrangers

Collaboratif : il permet les échanges entre les partenaires d'un
même projet

Sécurisé : hébergé sur un serveur à l'INIST-CNRS en France

DMP OPIDoR



Créer / rédiger un PGD à partir d'un modèle selon le financeur (ANR, Commission Européenne, Science Europe, INCa (Institut national du Cancer) ...) ou organisme de recherche (INRAE, Cirad ...)

Interactif : une assistance conseil auprès des services d'appui

Télécharger le PGD sous différents formats (pdf / doc ...)

Attention : DMP OPIDoR n'est pas un archive de PGD

DMPs : trouver des exemples



DMP publics

Les DMPs publics sont créés à l'aide de DMP OPIDoR et partagés publiquement par leurs propriétaires. Ils n'ont pas été vérifiés pour leur qualité, leur exhaustivité ou leur adhésion aux lignes directrices des financeurs.

Titre du projet	Modèle	Organisme	Propriétaire	Dernière mise à jour	Télécharger
DMP du projet "ANR WISPER-Water and Ice-related thermo-mechanical processes in the fractures of Steep alpine bedrock Permafrost"	ANR - Modèle de PGD (français)	CNRS	Florence Magnin	22/11/2022	
DMP du projet "Plan de Paradis's Projet"	Science Europe : modèle structuré	Université de Paris 8 - Vincennes-Saint-Denis	Paradis KAHOU	22/11/2022	
DMP de "SPRAS - the automatic annotation and analyses of speech"	Research Software Management Plan template (PRESOFT project)	CNRS	Brigitte Bigi	21/11/2022	
DMP du projet "PGD 1 : Suivi (factif) de population de poissons dans le lac du Bourget"	Science Europe : modèle structuré	Sorbonne Université	Merig AKDOGAN	17/11/2022	
DMP du projet "PGD Unité Expérimentale Horticole"	INRAE - Trame Structure	INRAE - Institut national de recherche pour l'agriculture l'alimentation et l'environnement	David LANOUE	17/11/2022	
DMP du projet "Plateforme de Probiotique Clinique"	INRAE - Trame Structure	Université de Montpellier	Jerome vialaret	15/11/2022	
DMP du projet "Les vinyles : objets de nostalgie et objets de collection"	DMP - Université Paris 8 Vincennes-Saint-Denis (français)	Université de Paris 8 - Vincennes-Saint-Denis	Aurélié Sévère	15/11/2022	



Home

Public DMPs

Public Dataset Desc.

Co-Branding

Support

Send feedback

Start new DMP

All (10) Public DMPs (5) Public Datasets (5)

Sort by: Published

Search

DMP

Published: 31 October 2022

Energy Optimization, Inefficiency Detection and Decarbonization using AI (EnergAI)

Owner: Published - Version 0 - Grant: European Federation of Data Driven Innovation Hubs

Contained Datasets: (3)

Cascais Municipality Office Energy Dataset, Energy Consumption (Gas and Electricity) - Dublin Civic Offices 2009-2012 DCC, REFIT: Electrical Load Measurements (Cleaned)

Export All DMP Versions

Dataset

Published: 31 October 2022

Energy Consumption (Gas and Electricity) - Dublin Civic Offices 2009-2012 DCC

Owner: Public - Grant: European Federation of Data Driven Innovation Hubs

Part of: DMP Energy Optimization, Inefficiency Detection and Decarbonization using AI (EnergAI)

Export

This website uses cookies to enhance your user experience

Bienvenue !

DMP OPIDoR vous accompagne à travers l'élaboration et la mise en pratique de plans de gestion de données et de logiciels.



Accessible à la communauté scientifique de l'ESR et à ses partenaires français ou étrangers



Personnalisable par tout organisme de recherche pour la mise en place de sa politique de données



Enrichi par des exemples et des recommandations adaptés à l'environnement de recherche



Collaboratif pour permettre les échanges entre les partenaires d'un même projet et les services d'accompagnement



Machine actionnable pour faciliter la saisie et les interactions avec les services impliqués dans la gestion des données

DMP OPIDoR évolue grâce à vos retours. Les développements s'inscrivent dans le cadre d'une collaboration internationale autour du logiciel open source DMPRoadmap

Planifiez et adoptez les outils pour Optimiser le Partage et l'Interopérabilité des Données de la Recherche

[Découvrez DMP OPIDoR](#)

Se connecter

Créer un compte

Accès institutionnel

Se connecter

Accès individuel

* Courriel

gemma.davies@univ-brest.fr

* Mot de passe

••••••••

[Mot de passe oublié ?](#)

Se souvenir de l'adresse

Se connecter

PGD : questions à se poser

- A qui appartiennent les données, qui est responsable de leur gestion?
- Quel espace sécurisé de stockage pour mes données?
 - Sauvegarde selon la règle 3-2-1 : 3 copies sur 2 supports différents dont 1 à distance
 - Volumétrie et nature des données va influencer sur le lieu de stockage
- Quels standards, normes en usage dans ma discipline, mon laboratoire?
- Est-ce qu'il y a besoin d'une méthode ou d'un outil logiciel spécifique pour accéder à mes données?
- Est-ce pertinent de tout partager? Déterminer degré d'ouverture. Pour quel public? Limites d'usage? Restrictions au partage? Quand partager? Dois-je me mettre en conformité avec une législation? Laquelle?
- Est-ce pertinent de tout conserver? Quelle durée de conservation?
 - Outre les données brutes, faut-il partager les différentes strates d'interprétations?

Réfléchir au devenir de vos données et les gérer, une habitude à prendre dès maintenant!

- Pour aller plus loin
 - [Dorandum](#) pour s'auto-former
 - Suivre [l'actualité de la science ouverte sur le site du COSO](#)
 - [Le guide Science Europe pour rédiger un PGD](#) + choisir un entrepôt + modèle DMP structuré sur DMPOpidor
 - Livre de Christine L. Borgman accessible sur OpenEdition, [Qu'est-ce que le travail scientifique des données?](#)
 - [Exemple rédaction data paper en SHS](#)
 - Retrouver des informations et des liens utiles sur notre page [Ubodoc consacrée aux Données de la recherche](#)