

Science ouverte : PGD/DMP et bailleurs de fonds

Centre d'Economie de la Sorbonne (CES - UMR 8174), 17 juin 2021

Amélie Collin, Centre de documentation du CES



Science Ouverte

« La science ouverte est la diffusion sans entrave des publications et des données de la recherche. Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et -autant que possible- aux données de la recherche. »

Source https://www.ouvrirlascience.fr/category/science_ouverte/

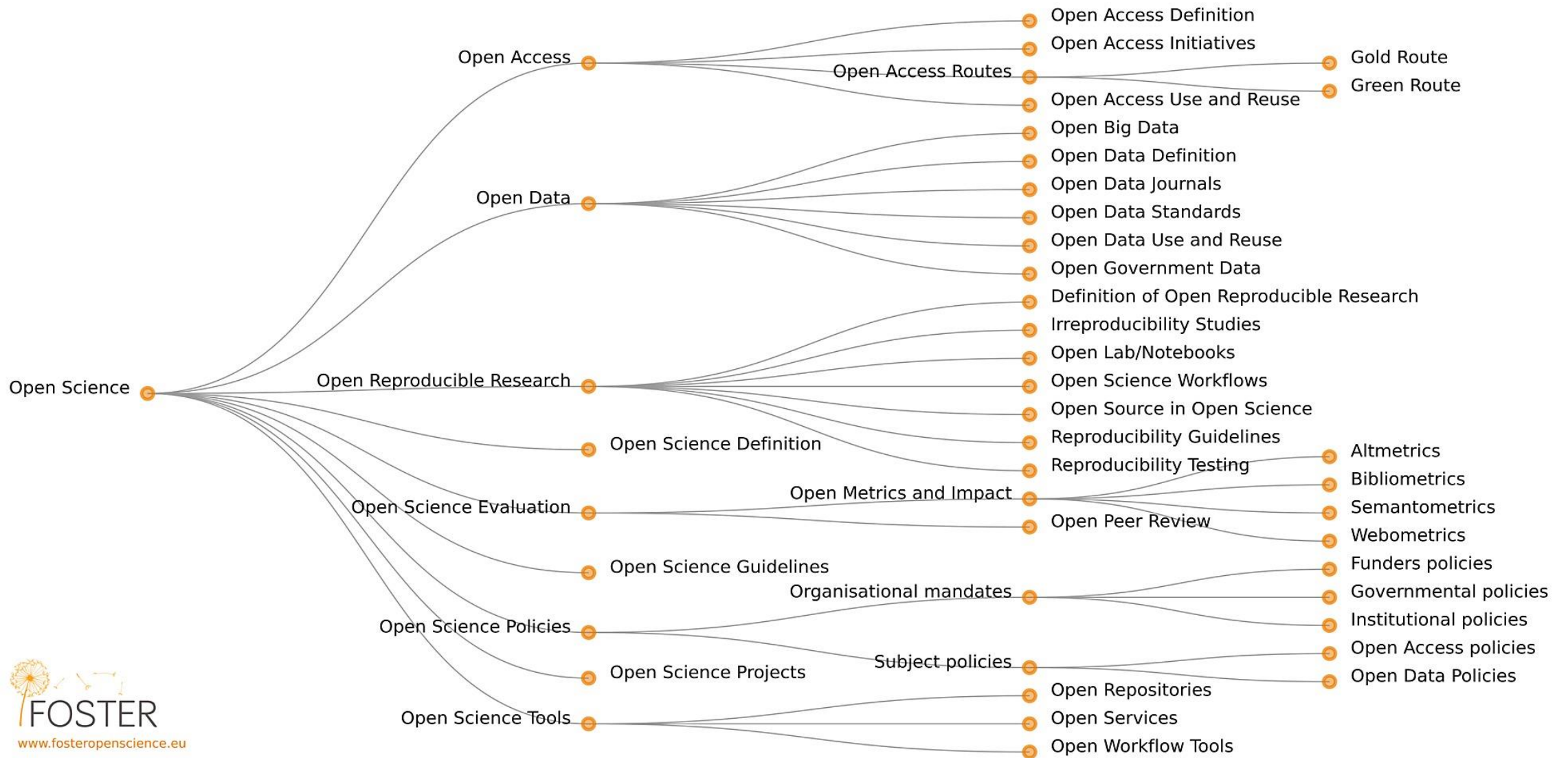
Le CNRS a diffusé en novembre 2020 son plan [Données de la Recherche](#), dans la lignée de sa feuille de route pour la science ouverte de 2019.

D'autres établissements d'enseignement supérieur et recherche s'emparent eux aussi de missions en lien avec la science ouverte, ou ont déjà mis en œuvre des actions en ce sens.

Objectifs du PNSO : 3 axes et 9 mesures

- Généraliser l'accès ouvert aux publications scientifiques
 - 1 - Rendre obligatoire la publication en accès ouvert des articles et livres issus de recherches financées par appel d'offres sur fonds publics.
 - 2 - Créer un fond pour la science ouverte.
 - 3 - Soutenir l'archive ouverte nationale HAL et simplifier le dépôt par les chercheurs qui publient en accès ouvert sur d'autres plateformes dans le monde.
- Structurer et ouvrir les données de recherche, le code source, les logiciels
 - 4 - Rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics.
 - 5 - Créer la fonction d'administrateur des données et le réseau associé au sein des établissements.
 - 6 - Créer les conditions et promouvoir l'adoption d'une politique de données ouvertes associées aux articles publiés par les chercheurs.
- S'inscrire dans une dynamique durable, européenne et internationale
 - 7 - Développer les compétences en matière de science ouverte notamment au sein des écoles doctorales.
 - 8 - Engager les opérateurs de la recherche à se doter d'une politique de science ouverte.
 - 9 - Contribuer activement à la structuration européenne au sein du European Open Science Cloud et par la participation à GO FAIR.

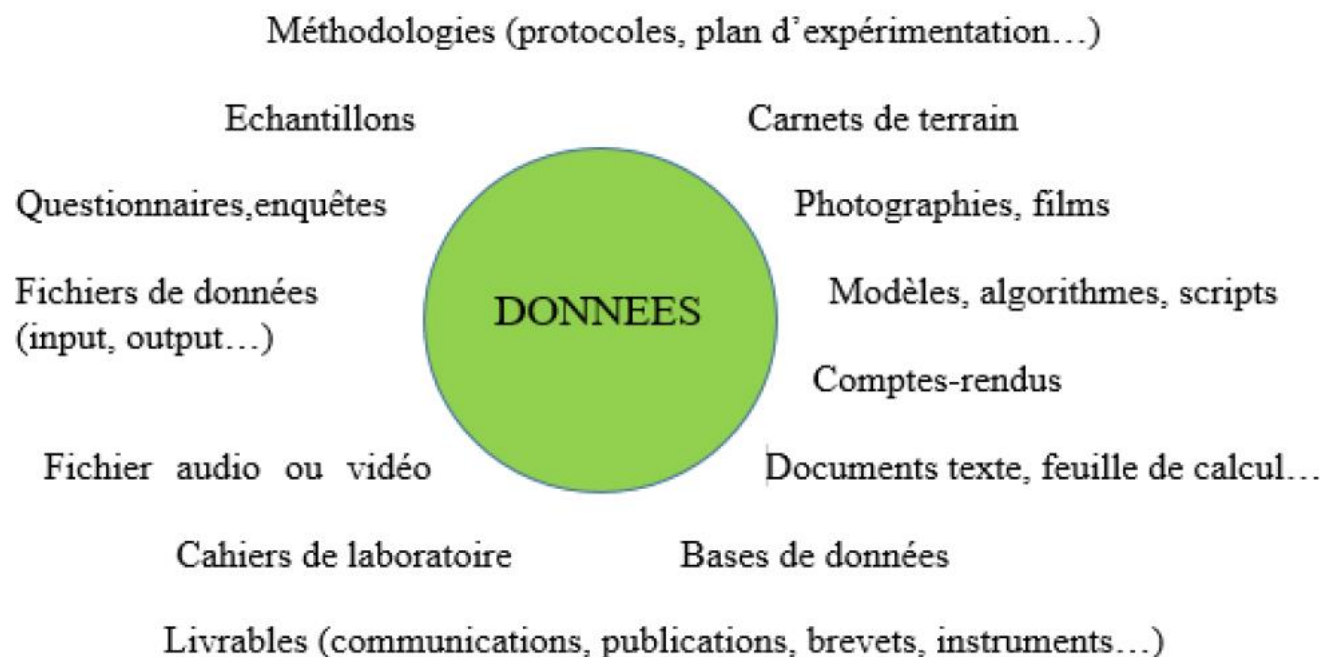
Open Science Taxonomy



Données de la recherche

« 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 des résultats de recherche. »

OCDE, *Principes et lignes directrices de l'OCDE, pour l'accès aux données de la recherche financée sur fonds publics*, 2007
<https://www.oecd.org/fr/science/inno/38500823.pdf>



Données de la recherche : enjeux



- **Pourquoi s'intéresser à la gestion des DR ?**

Enjeux d'utilisation des données :

- Croisements fertiles issus d'ensembles hétérogènes (analyse des données)
- La donnée comme guide de la recherche (*data driven science*)
- La donnée comme validation de la recherche

Enjeux de conservation des données :

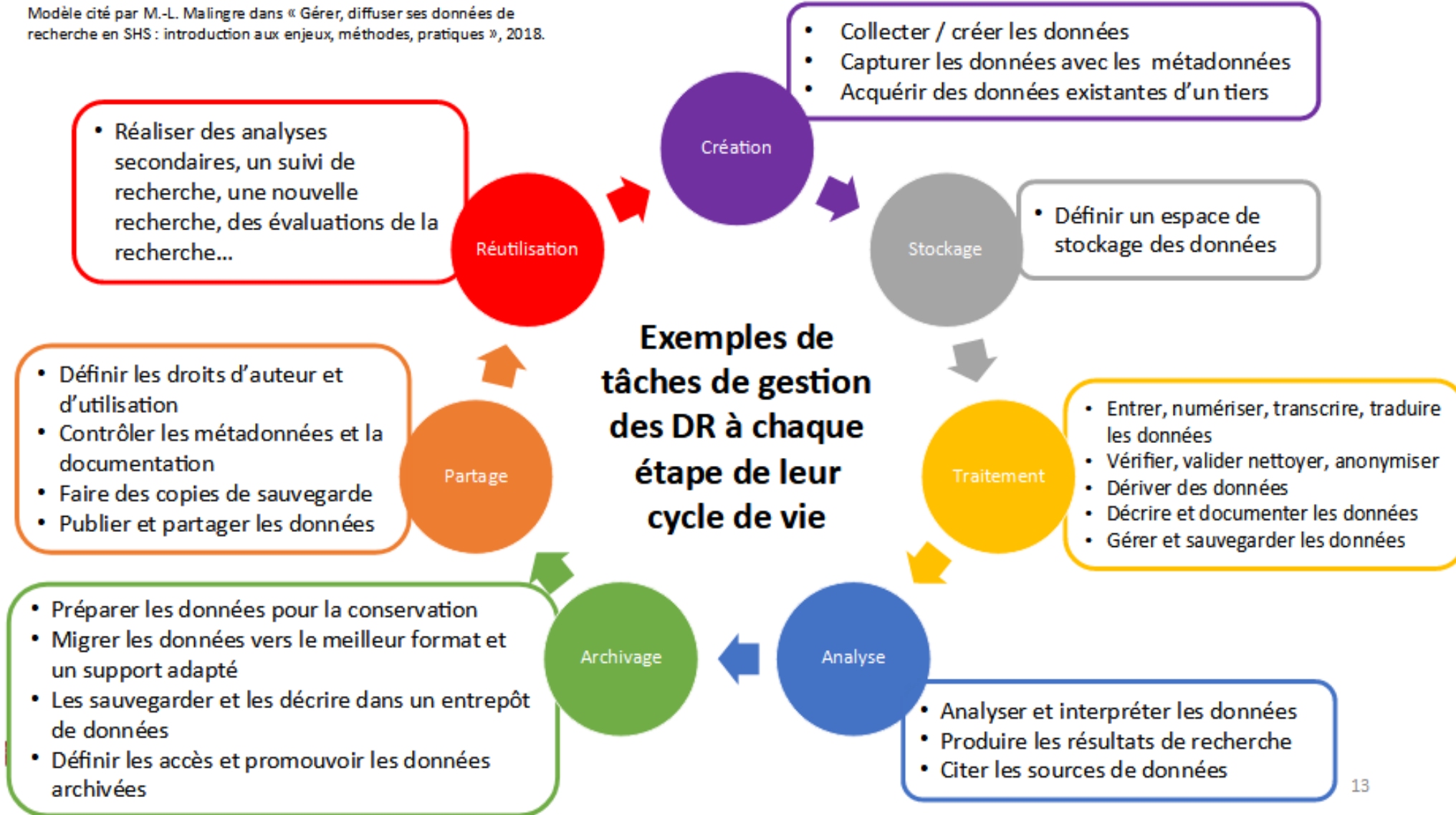
- Les difficultés, nombreuses négligences
- Eviter la perte de données uniques, positives, négatives, riches en informations

Enjeux de diffusion des données :

- Dépôt dans un entrepôt
- Matériel d'accompagnement d'une publication
- Exhumation de données « fossilisées » (les publications permettent d'accéder à environ 10 % des données, le reste demeurant disponible mais non utilisé sur les disques durs d'ordinateurs)

Gestion des données de la recherche : cycle de vie

Modèle cité par M.-L. Malingre dans « Gérer, diffuser ses données de recherche en SHS : introduction aux enjeux, méthodes, pratiques », 2018.



Gestion des données de la recherche : entrepôts de stockage et diffusion

- **Plateformes d'archivage des données**

Institutionnels (en France) :

[Dataverse Cirad](#), [Datapartage](#) (INRAE), [DataSuds](#) (IRD)

Pluridisciplinaires et internationaux :

[Zenodo](#), [Dryad](#), [Figshare](#)

Thématiques ou disciplinaires :

[Nakala](#) (Huma-Num, SHS), [GenBank](#) (séquences génétiques), [TRY](#) (caractères botaniques), [GBIF](#) (biodiversité), [Pangaea](#) (sciences de la terre et de l'environnement), [Movebank](#) (mobilité animale), [West African Vegetation](#), [DataFirst](#) (enquêtes socio-économiques en Afrique), [Protocols.io](#) (protocoles), etc.

Liés à un éditeur : [GigaDB](#) (Oxford Univ. Press), [Dataverse Ubiquity Press](#), [Dataverse Economics](#) (Harvard Library)

- **Annuaire d'entrepôts d'archivage des données**

[CoreTrustSeal](#) : annuaire d'entrepôts de données certifiés - recherche par pays via une carte interactive (ex. France = CINES).

[OAD \[Open Access Directory\] - data repositories](#) : liste par discipline des entrepôts et des bases de données pour les données ouvertes

[OpenDOAR](#) : répertoire mondial des entrepôts pour les données ouvertes et leurs politiques.

[Re3data.ORG - register of research data repositories](#) : moteur de recherche - description détaillée de plus de 2000 entrepôts de données.

Gestion des données de la recherche : entrepôts de stockage

- **Comment choisir un entrepôt de données ?**

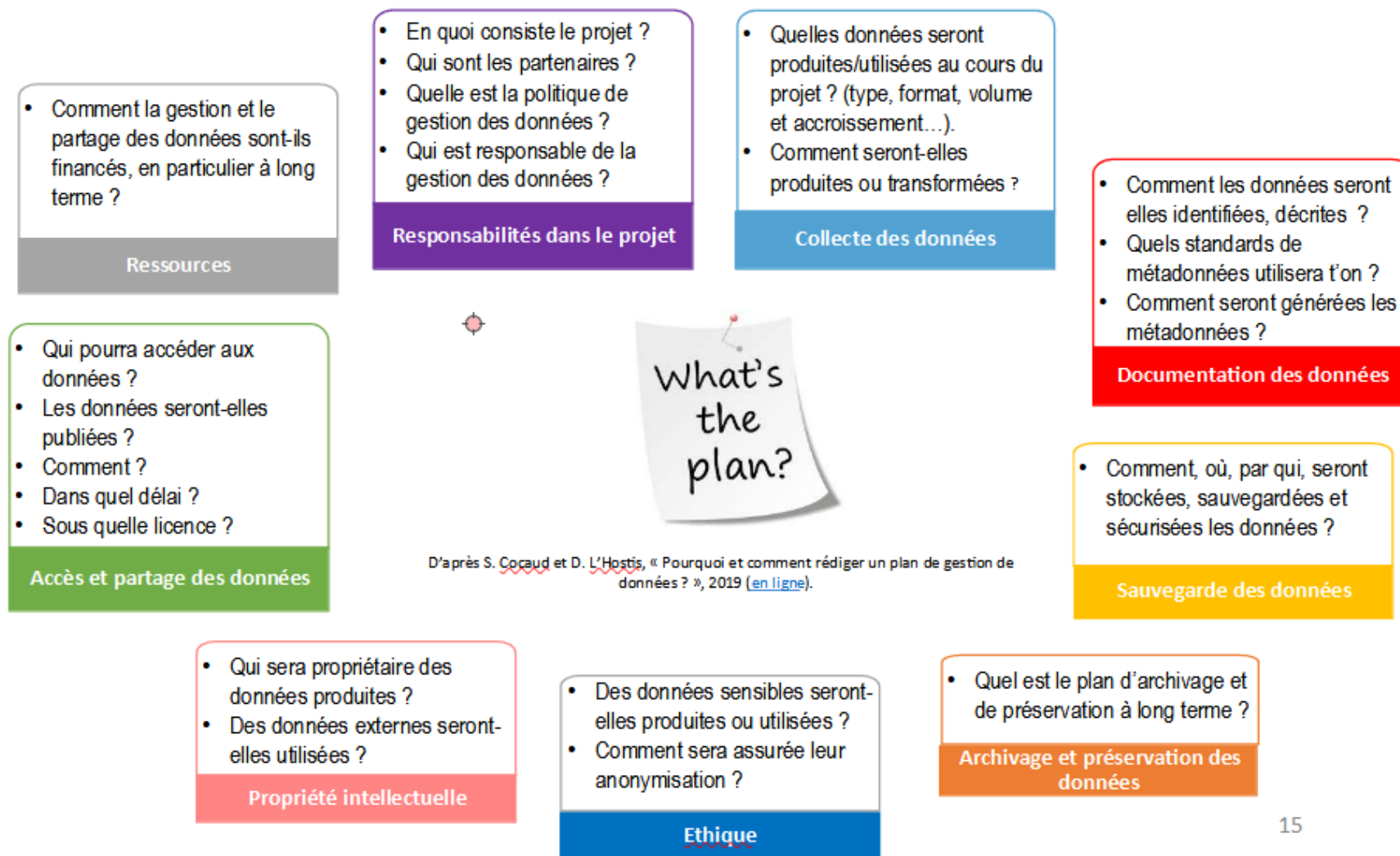
Re3Data : répertoire d'entrepôts de données créé en 2012 par le consortium international DataCite qui œuvre pour le partage des données de recherche. Re3data renseigne sur plus de 2450 entrepôts et sur les licences proposées par chacun.

Repository Finder : outil permettant de trouver un entrepôt de données. Cet outil, hébergé par DataCite, cherche des entrepôts dans le répertoire Re3data.

Fairsharing : guide sur les normes, standards, entrepôts et recommandations en termes de données et métadonnées.

Cat OPIDoR : catalogue des services français dédiés aux données scientifiques, hébergé par le CNRS. Cat OPIDoR présente une liste de 58 entrepôts de données en France.

Le Plan de gestion de données (PGD) ou Data Management Plan (DMP)



Quelques exemples de PGD

Un PGD de [thèse publié sous forme d'article de revue](#)

Un PGD RCUK Research Councils rédigé [sous forme de tableau](#)

Un PGD ERC [très court](#) (en cours, 4 p., v1, Contract No: [833438](#))

Un PGD HORIZON 2020 intégrant [figures et tableaux](#) (clos, 30 p., Contract No: [692819](#))

Un PGD HORIZON 2020 [avec des annexes](#) (en cours, 54 p., Contract No: [815001](#))

Un PGD de thèse avec [stratégie de sauvegarde](#)

Un PGD avec accent sur les [aspects éthiques et juridiques](#)

[OpenAIRE](#) a constitué une collection actuellement de 841 PGD relatifs à des projets Horizon 2020. Présentation de la collection et lien vers l'entrepôt :
<https://www.openaire.eu/blogs/establishing-a-collection-of-841-horizon-2020-data-management-plans>

Le Plan de gestion de données (PGD) : quels intérêts ?

Pour son propre intérêt

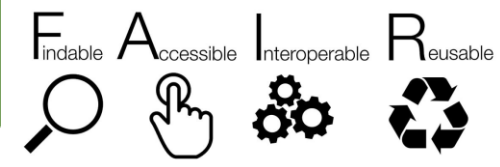
Pour répondre à des enjeux politiques

Pour répondre à des enjeux financiers

Pour répondre aux exigences des financeurs



Le Plan de gestion de données (PGD) : les principes FAIR



Facile à trouver

Findable 	Persistent Identifiers (PIDs) 	Rich metadata 	Indexed data repositories 	PIDs in metadata
---------------------	--	--------------------------	--------------------------------------	-----------------------------

Accessible

Accessible 	Standard communications protocol 	Open, free protocol 	Authentication, where necessary 	Metadata is always available
-----------------------	---	--------------------------------	--	---

Interopérable

Interoperable 	Vocabularies 	Vocabularies are FAIR 	Linked metadata
--------------------------	-------------------------	----------------------------------	----------------------------

Réutilisable

Reusable 	Metadata have multiple attributes 	Usage license 	Provenance 	Community standards
---------------------	--	--------------------------	-----------------------	--------------------------------



<https://www.g-fair.org/>

Source tableau : Australian National Data Service <https://www.ands.org.au/>

Logo GoFAIR : <https://www.go-fair.org/>

Le Plan de gestion de données (PGD) : les *Digital Object Identifier* (DOI)

Les PGD peuvent obtenir des DOI, grâce à une révision du schéma de métadonnées de [DataCite](#) qui inclut désormais la valeur "OutputsManagementPlan" dans son vocabulaire contrôlé pour le champ "resourceTypeGeneral".

⇒ *machine-actionable Data Management Plan* (maDMP)

Ressources complémentaires :

- K. Garza et M. Buys, « A Brave New PID: DMP-IDs », *DataCite Blog*, 2021 (en ligne <https://blog.datacite.org/announcing-dmp-ids/>).
- DORANum - [Petit guide DOI du chercheur](#)
- Opidor - [Attribuer un identifiant pérenne aux données de la recherche](#)
- [#activeDMPs](#) : site qui réunit les travaux en cours sur les maDMP

Le PGD ANR : modèle PNSO



05/09/2019
L'ANR met en place un plan de gestion des données pour les projets financés dès 2019



Dans le cadre de sa politique science ouverte, l'Agence nationale de la recherche (ANR) demande l'élaboration d'un Plan de Gestion des Données (PGD) pour les projets financés à partir de 2019. Ce document vise à décrire la façon dont les données de recherche seront produites, collectées, réutilisées, stockées, protégées, disséminées ou encore conservées à long terme. Objectif : une meilleure anticipation

[Source de l'article](#)



[Plan d'action 2019 de l'ANR](#), p. 9
Voir aussi les [recommandations du CoSO](#)

Plan d'action 2019

Date de publication :
28 juillet 2018

Ce document a été adopté le 02 juillet 2018 par le Conseil d'administration de l'ANR



Plan d'action 2020

Appel à projets générique 2020

(version 1.1)

Date de publication 25 juillet 2019



C.4. Publications scientifiques et données de la recherche

Dans le cadre de la contribution de l'ANR à la promotion et à la mise en œuvre de la science ouverte, et en lien avec le plan national pour la science ouverte, les partenaires au projet déposé s'engagent en cas de financement :

- à déposer leurs publications scientifiques (texte intégral) issues du projet de recherche, de développement et d'innovation dans une archive ouverte, soit directement dans HAL soit par l'intermédiaire d'une archive institutionnelle locale, dans les conditions de l'article 30 de la Loi « Pour une République numérique »^{20,21}
- à fournir dans les 6 mois qui suivent le démarrage du projet un plan de gestion des données (PGD)²² selon des modalités communiquées dans les Conditions particulières et le Règlement financier de l'ANR.

Par ailleurs, l'ANR recommande de privilégier la publication dans des revues ou ouvrages nativement en accès ouvert.²³

« Le coordinateur ou la coordinatrice s'engage [...] à fournir un plan de gestion des données (DMP) actualisé tout au long de la vie du projet selon les modalités communiquées lors du conventionnement. » ([Guide de l'AAPG 2020](#), p. 20)

Le PGD ANR : modèle PNSO

Ressources PGD ANR :

- ▶ [Politique en matière d'éthique et d'intégrité scientifique de l'ANR \(2019\)](#)
- ▶ [Modèle de plan de gestion des données ANR \(2019\)](#)
- ▶ DORANum - [Guide de relecture de PGD - Modèle ANR](#)
- ▶ Webinaire - [La politique science ouverte de l'ANR et le Plan de Gestion des Données \(PGD\), 14 novembre 2019](#)

Le PGD Horizon 2020 : modèle FAIR DMP



H2020 Programme

AGA – Annotated Model Grant Agreement

Version 5.2
26 June 2019

Disclaimer:
This guide is aimed at assisting beneficiaries. It is provided for information purposes only and is not intended to replace consultation of any applicable legal sources. Neither the Commission nor the Executive Agencies (or any person acting on their behalf) can be held responsible for the use made of this guidance document.

ARTICLE 29 — DISSEMINATION OF RESULTS — OPEN ACCESS — VISIBILITY OF EU FUNDING

1. Dissemination of results

Unless it goes against their legitimate interests, the beneficiaries must — as soon as possible (but not before a decision on their possible protection) — **disseminate** their results (i.e. make them public).

Results that are disclosed too early (i.e. before the decision on their protection) run the risk being invalidated.

2. Open access to scientific publications

What? Beneficiaries must ensure **open, free-of-charge access** to the end-user to **peer-reviewed scientific publications** relating to their results.

'Peer-reviewed publications' means publications that have been evaluated by other scholars (e.g. articles in scientific journals).

3. Open access to research data (Extended Open Research Data Pilot)

What? Beneficiaries of actions that participate in the Open Research Data Pilot (ORDP) must:

- develop a **data management plan (DMP)** as a deliverable of the project and
- give **open, free-of-charge access** to the end-user to **digital research data** generated during the action.

The purpose of the ORDP is therefore twofold: research data management and open access.

How? Open access to digital research data involves 3 actions:

- 1) Develop a **data management plan (DMP)** by month six of the project as a deliverable. The DMP is a 'living document' to be revised as appropriate in the duration of the project

Le PGD Horizon 2020 : modèle FAIR DMP

Ressources DMP Horizon 2020 :

- ▶ [H2020 Open access & data management](#)
- ▶ [Guidelines on FAIR Data Managemnt in Horizon 2020](#) par la C.E. (v3, 2016) /
- ▶ [Lignes directrices pour la gestion des données de recherche dans H2020](#) par la C.E. (v3, 2016, trad. Fr Inist 2016)
- ▶ [Modèle de plan de gestion des données Horizon 2020](#) (*Data Management Plan - DMP*)
- ▶ Modèle DMP Horizon 2020 annoté : [AGA – Annotated Model Grant Agreement](#) (2019)
- ▶ [Guide d'aide à la rédaction d'un plan de gestion des données](#) de Sorbonne Paris Cité (2015)
- ▶ [Fiche P.C.N. sur le libre accès aux données de recherche](#)
- ▶ [DMPonline](#) développée conjointement par le Digital Curation Center (DCC) et le University of California Curation Center (UC3).

Résumé des outils et ressources

Outils génériques

- ▶ [Cat OPIDOR](#) : wiki des services dédiés aux données de la recherche
- ▶ [DATAVERSE](#) : application web pour partager, archiver, retrouver et citer des données de la recherche
- ▶ [DORANum](#) : plateforme nationale de DONnées de la Recherche Apprentissage Numérique (Inist/CNRS)
- ▶ [OPIDOR](#) : portail national de l'offre de services dédiés aux données de recherche (Inist/CNRS)
- ▶ [Research Data Management Toolkit](#) : boîte à outils pour gérer les données de recherche

Outils DMP

- ▶ [Questionnaire d'aide à la décision pour diffuser ses données de recherche](#), Mathieu Andro, Nathalie Morcrette et Nathalie Grandon (Inra)
- ▶ [DMP OPIDoR](#) : rédaction collaborative de PGD suivant différents modèles (Inist/CNRS)

Outils Métadonnées

- ▶ [DataCite Metadata Generator](#) : outil pour générer des métadonnées (XML)
- ▶ [Metadata Anonymisation Toolkit](#) (MAT) : outil pour nettoyer les métadonnées des fichiers avant de les publier
- ▶ Le [guide OTELo](#) : guide de bonnes pratiques pour la gestion et la valorisation des données de la recherche (Inist/CNRS)
- ▶ [Catalogue de schémas de métadonnées](#) élaboré par l'un des groupes de travail de la Research Data Alliance (RDA) .

Ressources pour aller plus loin

Ouvrir la science

Science ouverte au CNRS

Science ouverte au MESRI

La Gestion des données

- DORANum – Guides d’auto-formation à la gestion des données
- Guide CoopIST – Gérer les données de la recherche
- Open science : comment l’inscrire dans ses pratiques de recherche ? / Mariannig LE BECHEC, *Journée thématique - CALMIP / GRICAD* (2019)
- Ouvrir la science – Guide pratique pour une harmonisation internationale de la gestion des données de la recherche (2019)

Aspects juridiques et éthiques

- InSHS – Guide pour la recherche : Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte (v2, février 2021)
- DORANum – Aspects juridiques, éthiques, intégrité scientifique
- La diffusion numérique des données en SHS – Guide de bonnes pratiques éthiques et juridiques de l’AMU

Les entrepôts de données de recherche

- Sylvie Cocard, Pascal Aventurier. Les entrepôts de données de recherche. *Participer à l’organisation du management des données de la recherche, gestion de contenu et documentation des données. Action Nationale de Formation organisée par les réseaux Renatis et Médiçi*, Centre National de la Recherche Scientifique (CNRS). FRA., Jul 2017, Vandoeuvre-les-Nancy, France, pp.63
<10.15454/1.4993537478868977E12>. <hal-01595599>

Principes FAIR

- Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez, Réaliser un plan de gestion de données « FAIR » : modèle, 2018. <sic 01690547v2>
- Guide INRAE - Produire des données FAIR
- Guide Institut Pasteur – Comment rendre ses données FAIR ? (2020)

Ressources pour aller plus loin

Le PGD

- Guide CoopIST – [Se familiariser avec des plans de gestion de données de la recherche](#) / Marie-Claude Deboin (2018)
- [CNRS – Plan de gestion de données de recherche](#) (novembre 2020)
- DORANum – [Plan de gestion de données. Pourquoi et comment rédiger un plan de gestion de données ?](#)
- [DMP OPIDoR](#), outil collaboratif d'aide à la rédaction en ligne de DMP
- [DMP OPIDoR, modèles de DMP](#)

PGD et bailleurs de fonds

- [Les enjeux européens autour de la question des données. Les DMPs, H2020 et les programmes à venir](#) / Corinne Joffre, *Journée thématique - CALMIP / GRICAD* (2019)
- [Les DMPs concrètement. Pourquoi et comment rédiger un DMP dans le cadre d'un projet de recherche](#) / Chloée FABRE, *Journée thématique - CALMIP / GRICAD* (2019)

DOI pour PGD

- K. Garza et M. Buys, « A Brave New PID: DMP-IDs », *DataCite Blog*, 2021 (en ligne <https://blog.datacite.org/announcing-dmp-ids/>).
- DORANum - [Petit guide DOI du chercheur](#)
- Opidor - [Attribuer un identifiant pérenne aux données de la recherche](#)
- CoopIST – [Les principes du DOI](#)

Logiciels et code source

- Recommandations du groupe projet « [logiciels libres et open source](#) » du CoSO (2019).
- [Guide Institut Pasteur – Comment rendre son logiciel FAIR ?](#) (2021)

Data paper

- Guide CoopIST – [Rédiger et publier un data paper](#)
- [Du Plan de gestion de données au Data Paper](#) / Dominique L'HOSTIS, *Journée thématique - CALMIP / GRICAD* (2019)

Science ouverte : PGD / DMP et bailleurs de fonds

Centre d'Economie de la Sorbonne (CES - UMR 8174), 17 juin 2021

Merci de votre attention

Contacts :

CESDoc – amelie.collin@univ-paris1.fr et rachel.creppy@univ-paris1.fr

<https://cesdoc.pantheonsorbonne.fr/appui-recherche>

SCD université Panthéon-Sorbonne Paris 1 - appui-recherche-scd@univ-paris1.fr

IST CNRS - <https://www.science-ouverte.cnrs.fr/services-a-la-recherche/>

DDOR CNRS - <https://www.science-ouverte.cnrs.fr/ddor-cnrs-direction-des-donnees-ouvertes-de-la-recherche/>

