



Open Science Month

Introduction aux données de la recherche

Stela SUHAN (DiBISO)

5 octobre 2022

L'univers des données de la recherche vous semble obscur ? Venez découvrir ce que sont les données de la recherche, leur cycle de vie, les pratiques de leur gestion et les enjeux de leur ouverture



Plan

- Définition et classification des données de la recherche
- Cycle de vie des données
- Données FAIR
- Data Management Plan (DMP)
- Entrepôts de données
- Enjeux de l'ouverture des données de la recherche



Définir les données de la recherche

<https://fr.freepik.com/photos-vecteurs-libre/donnees>



**Matériel factuel, collecté,
observé, généré ou créé lors
d'un projet scientifique pour
valider des résultats de
recherche**

https://fr.freepik.com/vecteurs-libre/ensemble-scientifiques-faisant-recherche-banniere_5576173.htm

C'est quoi au juste les 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 les résultats de recherche** »

OCDE, « Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics », Paris, 2007, p.18

Peuvent être classées selon :

Leur nature ou forme

- Texte
- Numérique
- Multimédia

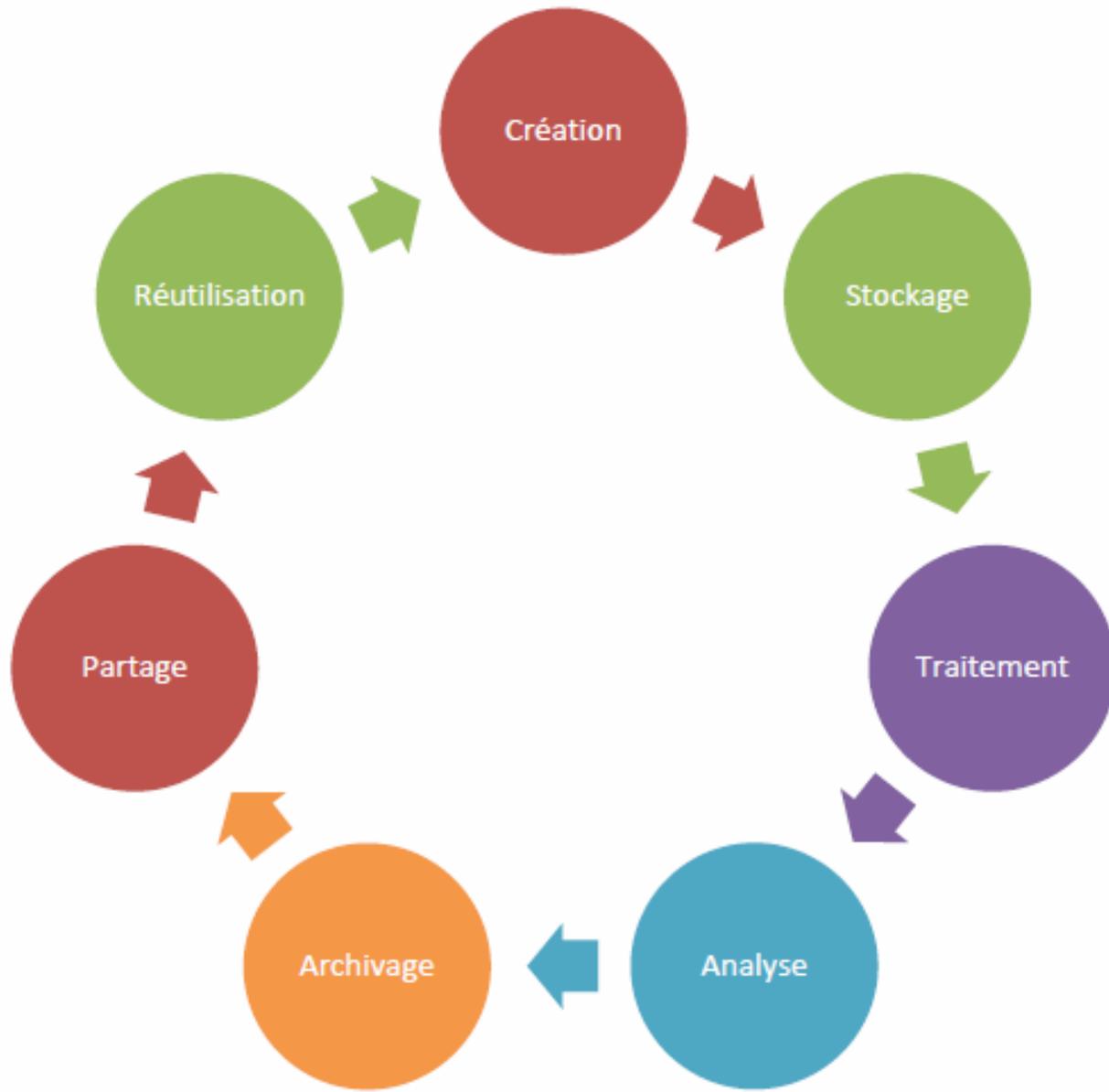
Niveau d'élaboration

- Données brutes
- Données traitées
- Données analysées

Définition d'un jeu de données (*Data Set*)

Un regroupement de données numériques, organisées dans un cadre structuré pour former un ensemble cohérent et pour répondre à une question de recherche.





Le cycle de vie des données

<https://data.ird.fr/gerer/>

Stockage, sauvegarde, archivage

- Règle du 3-2-1
- Gérer les versions
- Bonnes pratiques de nommage des fichiers
 - 30 caractères maximum
 - Noms des partenaires insérables si leur graphie est harmonisée entre les fichiers
 - Numéros de versions le cas échéant
 - Dates au format ISO : AAAA-MM-JJ
- Archivage pérenne se pose uniquement pour les données présentant une **valeur scientifique reconnue** par la communauté

Données FAIR

- **Faciles à trouver**

- Déposer les données dans un entrepôt
- Attribuer un identifiant unique et pérenne (DOI)
- Décrire les données par les métadonnées

- **Accessibles**

- Définir les conditions d'accès aux données
- Si possible, rendre les données accessibles librement
- En cas de d'accès restreint, rendre accessibles les métadonnées pour signaler l'existence des données

- **Interopérables**

- Formats ouverts
- Mises à disposition du code source du logiciel
- Standards des métadonnées et vocabulaires contrôlés

- **Réutilisables**

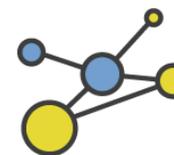
- Associer une licence de diffusion
- Décrire en détail les données, les contextualiser, les rendre compréhensibles



Findable



Accessible



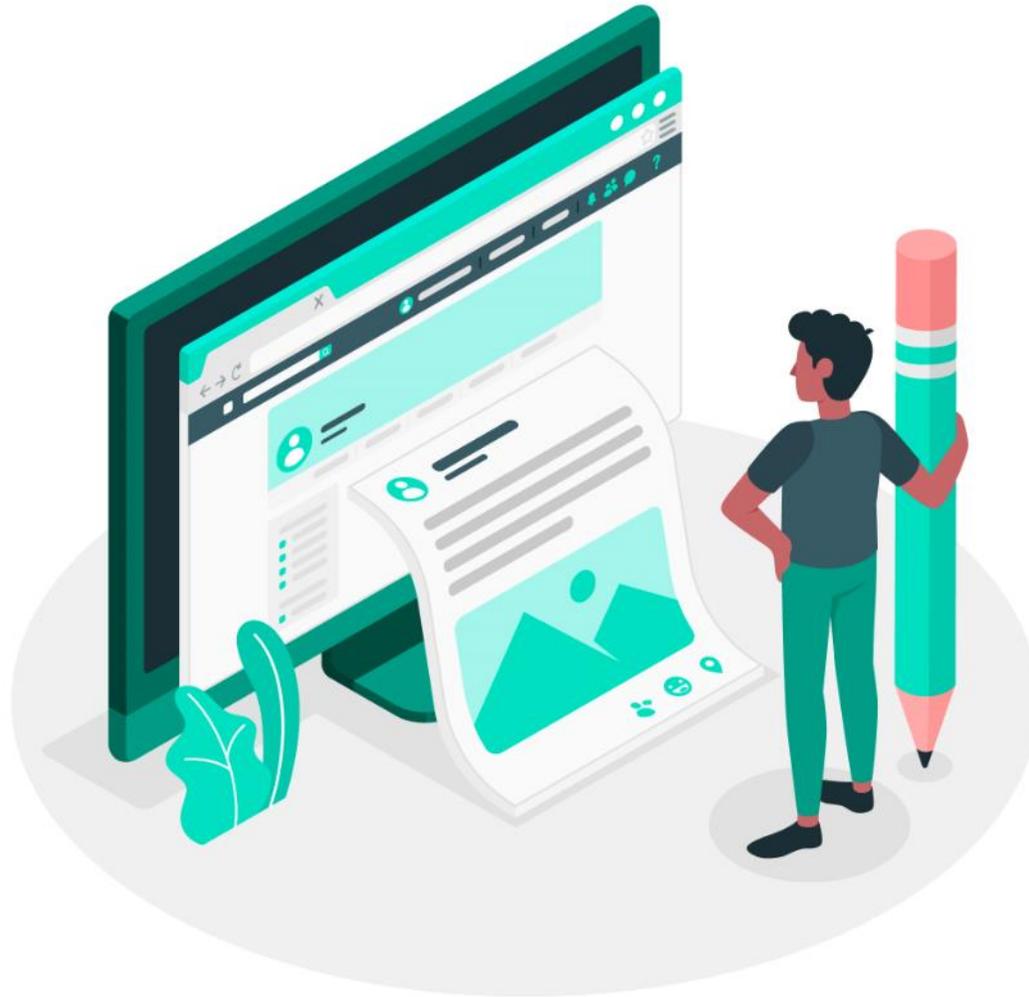
Interoperable



Reusable

<https://www.nlm.nih.gov/oet/ed/cde/tutorial/02-300.html>

Plan de Gestion des Données (PGD) ou Data Management Plan (DMP)



Les parties structurantes du DPM

- ❑ Description des données
- ❑ Documentation et qualité
- ❑ Exigences légales et éthiques
- ❑ Sauvegarde et stockage des données au cours du projet
- ❑ Stratégie de partage et d'ouverture des données
- ❑ Préservation à long terme des données
- ❑ Administration de la gestion du projet

Des outils d'aide à la rédaction

- <https://dmptool.org/> (Université de Californie)
- <https://argos.openaire.eu/home> (OpenAIRE)
- <https://dmponline.dcc.ac.uk/> (Digital Curation Centre)
- <https://dmp.opidor.fr/> (Inist - CNRS)
- OPIDoR: Optimiser le Partage et l'Interopérabilité des Données de la Recherche



Entrepôts des données (*Data repositories*)

- Re3data (*Registry of Research Data Repository*)
- FAIRsharing (*Sciences de la Vie*)
- CatOPIDoR (récence les services français)



Source image: https://fr.freepik.com/vecteurs-libre/gros-centre-donnees-rack-salle-serveurs-processus-ingenierie-travail-equipe-technologie-informatique-stockage-nuage_4103163.htm

Exemples des entrepôts

- **Recherche Data Gouv**
 - Nouvelle plateforme nationale fédérée des données de la recherche
- **Zenodo (Commission Européenne)**
 - entrepôt mixte (publications scientifiques et jeux de données), tous types de données
- **Figshare**
 - tous types de données
- **4TU.ResearchData**
 - entrepôt reconnu de données internationales pour la science, l'ingénierie et la conception
- **DRYAD**
 - données sous-jacentes aux publications
- **PANGAEA**
 - Entrepôt reconnu en sciences de la terre et de l'environnement
- **NAKALA**
 - entrepôt national français en SHS
- **UK Data Archive**
 - Données numériques en sciences sociales et humaines au Royaume-Uni



L'ouverture des données : un enjeu de science ouverte

<https://fr.freepik.com/photos-vecteurs-libre/affaires>

Ouverture des données impulsée par PNSO et Financeurs



- obligation de déposer les publications dans une archive ouverte
- obligation de rédaction d'un DMP
- **incitation à l'ouverture des données de la recherche**
- « **Partage précoce des connaissances et des données, et collaboration ouverte** », Horizon Europe (2021-2027)

Ouverture des données de la recherche



Mettre en libre accès les données issues d'un projet de recherche

Les données ouvertes sont accessibles à tous sans restriction

Les données partagées nécessitent une identification et autorisation des utilisateurs

Source: <https://www.universite-paris-saclay.fr/recherche/science-ouverte/les-donnees-de-la-recherche/ouvrir-ses-donnees-comment-ca-marche>

Benefits of Open Data



Researchers

- greater discoverability
- increased efficiency
- attracts funding & support
- new collaborations



Funders

- increased visibility & reuse of funded research
- greater funding impact
- greater ROI



General Public

- self-empowerment
- increased transparency
- greater engagement in science & research



Organisations/ NGOs

- enhanced access to research
- better information-sharing
- more effective advocacy/lobbying



National Governments

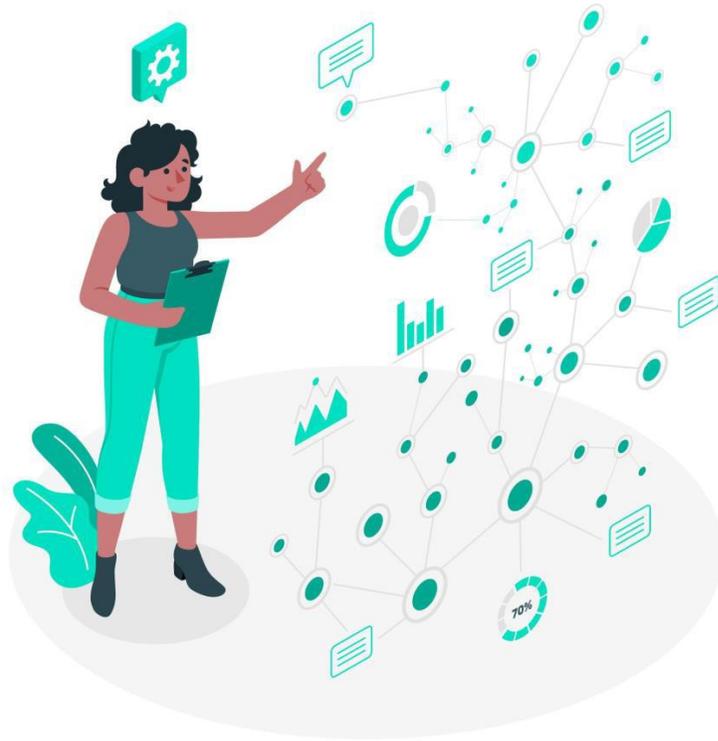
- data-driven decision making
- reduced government costs
- more effective & efficient government services

Source: <https://www.openaire.eu/what-is-open-research-data>

Data sharing by *NYU Health Sciences Library* (is about what happens when a researcher hasn't managed their data)



Available at: <https://www.youtube.com/watch?v=N2zK3sAtr-4>



<https://fr.freepik.com/photos-vecteurs-libre/graphique>

Pour aller plus loin...

- Page web de l'Université Paris Saclay
<https://www.universite-paris-saclay.fr/recherche/science-ouverte/les-donnees-de-la-recherche>
- La plateforme de formation DoRANum (Données de la Recherche: Apprentissage Numérique)
<https://doranum.fr/>
- Parcours interactif de DoRANum
<https://doranum.fr/enjeux-benefices/parcours-interactif-sur-la-gestion-des-donnees-de-la-recherche/>
- Jeux pédagogique de GopenDoRe en ligne
<https://doranum.fr/enjeux-benefices/gopendore-jeu-sur-la-gestion-des-donnees-de-recherche/>

N'hésitez pas à nous contacter !

Services proposés autour des données :

- Accompagnement à la diffusion des données
- Sensibilisation à la Fairisation des données
- Relecture de PGD
- Sensibilisation aux enjeux de la science ouverte

donnees-recherche@universite-paris-saclay.fr



designed by  freepik.com