



Service commun de la documentation

Département d'appui à la recherche

Atelier :

**« S'initier à la gestion des données
de la recherche »**

28/02/2022





Qu'entend-on par données de la recherche ?

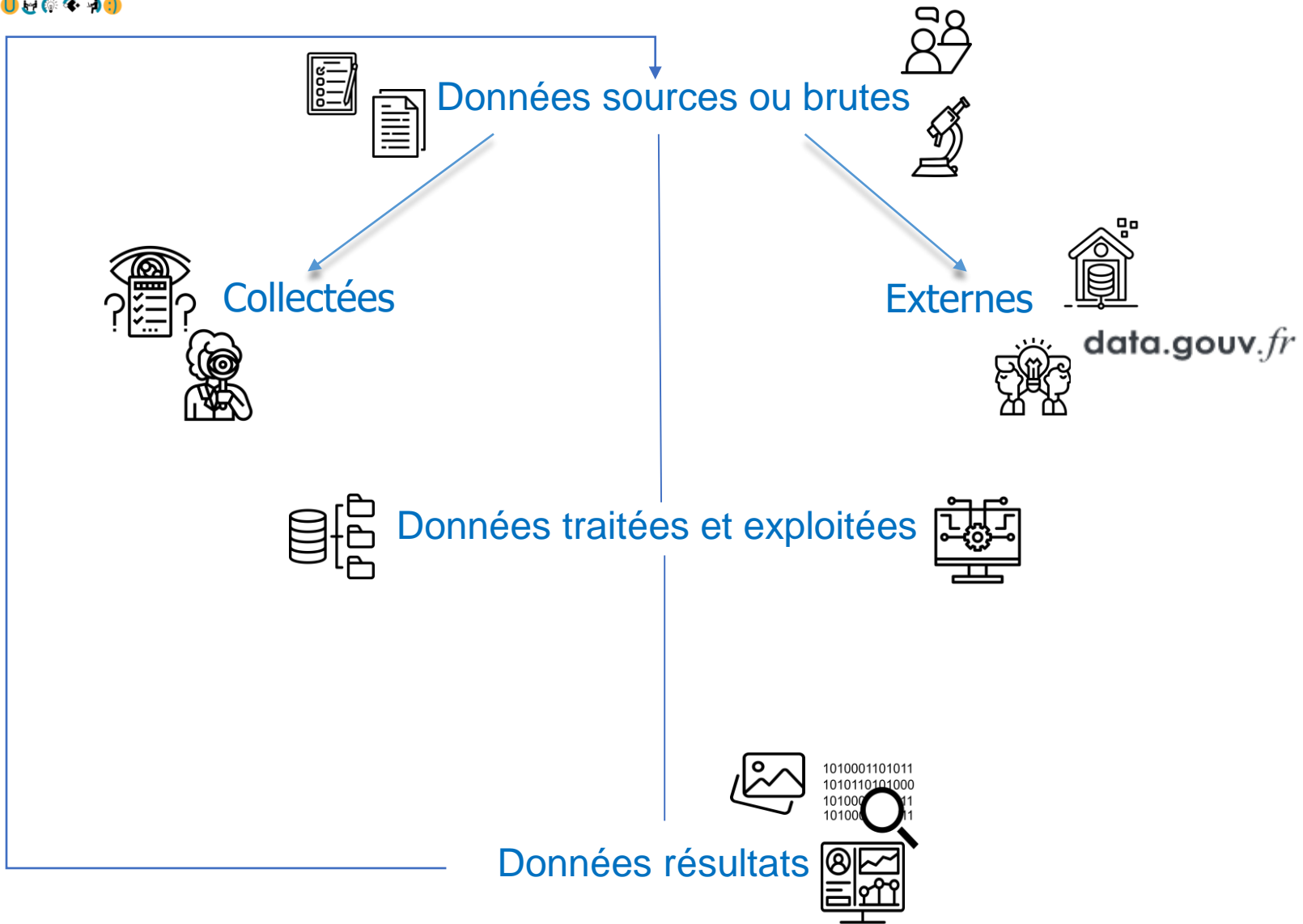




Définition des données de la recherche

Définition générale la plus utilisée, celle de l'OCDE (2007) ([Principes et lignes directrices pour l'accès aux données de la recherche financées sur fonds publics](#)) :

« Enregistrements factuels (chiffres, textes, images, sons) utilisés comme source principale pour la recherche scientifique et généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche »





Définition des données de la recherche

Données sources : Données qui constituent les matériaux de base de la recherche, avant tout traitement et analyse.

Données sources collectées : Données collectées et/ou compilées par la ou le chercheur(se) en direct ou à distance.

Données sources externes : Données non produites par la ou le chercheur(se) (*par exemple : corpus et archives textuelles collectées par d'autres personnes/institutions, statistiques produites par des acteurs publics/privés...*).

Données produites à l'issue des travaux de recherche après exploitation et traitement des données (*par exemple : synthèse, notes, photographies, tableaux, statistiques, modèles, codes, logiciels...*).



Quelques notions autour des données de la recherche

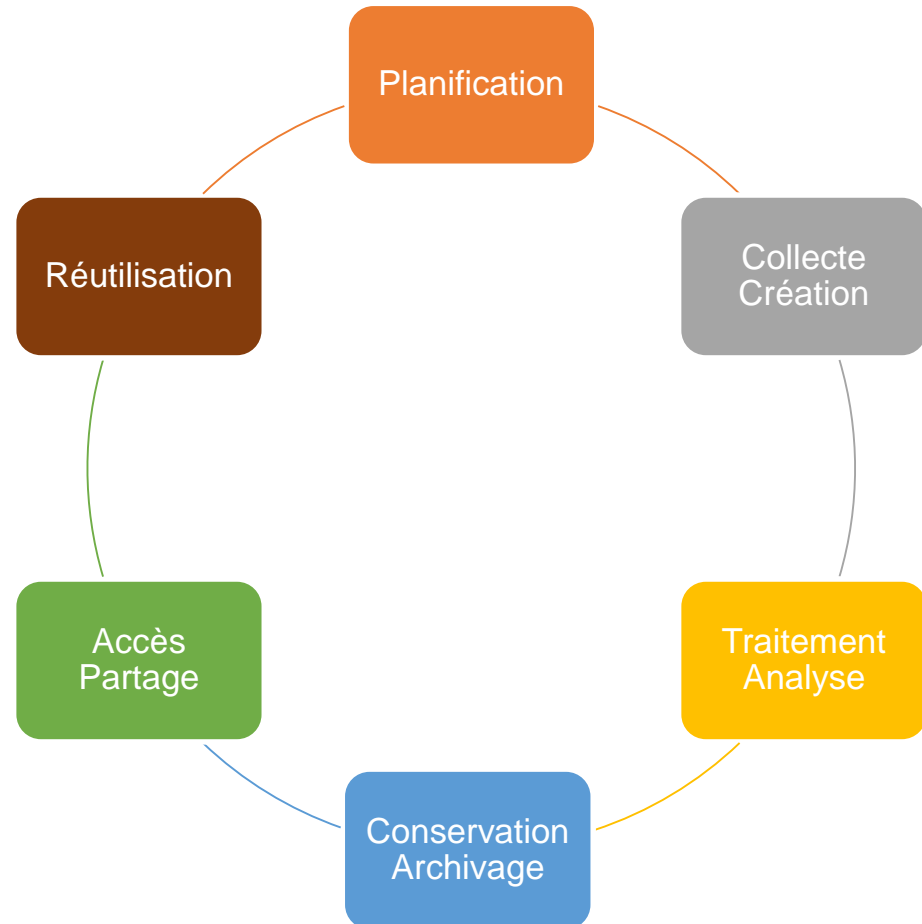




Le cycle de vie des données

Définition :

Ensemble des opérations réalisées tout au long du cycle de vie de la donnée pour en assurer la qualité et la pérennité.





Planification

- Concevoir et conceptualiser
- Mettre en place un Plan de Gestion de Données (PGD / DMP)

Collecte Création

- Préciser l'origine des données (collectées/ et ou créées)
- Décrire les données et métadonnées

Traitement Analyse

- Vérifier, nettoyer et valider les données
- Interpréter les données
- Gérer et stocker les données



Conservation Archivage

- Archiver les données à long terme
- Attribuer des identifiants pérennes (DOI)
- Se poser la question de la dimension juridique (licences)

Accès Partage

- Publier les articles et lier les data papers
- Appliquer une licence d'utilisation
- Choisir un entrepôt dans une logique de partage

Réutilisation

- Permettre de nouvelles recherches et la reproductibilité des résultats
- Réexaminer les résultats / données
- Enseigner et apprendre
- Respecter la licence de réutilisation



Les principes FAIR

Findable



Facile à trouver

- Descriptions
- Identifiant pérenne

Accessible



Facile à utiliser

- Protocole de communication ouvert
- Authentification
- Accès aux métadonnées

Interoperable



Facile à échanger

- Standards de description
- Format ouvert
- Données liées

Reusable



Facile à réutiliser

- Licence de réutilisation
- Provenance
- Format ouvert/standard



Le Plan de gestion de données

Outil essentiel dès le début des recherches, il expose comment seront gérées les données pendant et après le projet.

S'intéresse à l'ensemble du cycle de vie des données et est mis à jour et enrichi au fur et à mesure.

Sert à formaliser et partager ses pratiques sur la gestion des données durant l'ensemble de leur cycle de vie.

Document partagé avec toutes les parties prenantes du projet de recherche.

Livrable attendu par l'ANR et la commission européenne **6 mois après le début du projet.**

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





Le Plan de gestion de données

ANR : 6 rubriques

[Trame ANR](#)

1. Description des données.
2. Documentation et qualité des données.
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données et conservation à long terme
6. Responsabilités et ressources

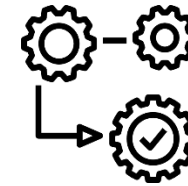
Commission européenne

[Trame Horizon Europe](#)

1. Data Summary
2. FAIR Data.
3. Other research outputs
4. Allocation of Resources
5. Data Security
6. Ethics
7. Other Issues



Apprendre à bien gérer ses données de recherche





Comment bien gérer ses données de recherche ?

Quelques pistes :

1. **Standardiser** sa collecte d'information (permet la reproductibilité des résultats).
2. **Organiser** dès le début du projet ses données de la recherche.
3. **Sauvegarder** et réaliser à intervalle régulier des copies de ses données (limite le risque de perte des données).
4. **Tracer et indexer** ses données, produire de l'information descriptive sur la source et le type des données (métadonnées, documentation...).



Standardiser la collecte d'information

Pourquoi ?

Si la collecte d'information n'est pas répétable la reproductibilité des résultats ne peut pas être faite.

Elisa, doctorante souhaite montrer un lien entre la délétion d'un gène chez la souris et la survenue de problème neurolocomoteurs. Elle a obtenu d'excellents résultats quelle présente en réunion d'équipe.

Son chef d'équipe lui demande le descriptif rédigé de son protocole. Elisa n'a pas pris le temps de rédiger quelque chose qui était dans sa tête ? Elle peut oublier son article.

Comment remédier à ce problème ?

En rédigeant **un protocole de collecte** des informations. Ce document va décrire la façon dont l'information est collectée et ce, donnée par donnée.



Organiser ses données

Pourquoi ?

- Permettre de retrouver et de protéger vos ressources.
- Gagner du temps dans les étapes de votre projet.

Comment ?

- En utilisant des standards de métadonnées adaptés au type de vos ressources.
- En réfléchissant à la manière de structurer vos données en fonction du traitement que vous allez réaliser.



Sauvegarder ses données

Pourquoi ?

Garantir la pérennité et l'intégrité de ses données de recherche.

Eviter d'altérer les informations recueillies.

Comment ?

Réaliser des sauvegardes régulières de ses données sur plusieurs supports sécurisés en appliquant dans l'idéal **la règle du 3-2-1** :

- garder 3 exemplaires des données,
- sur 2 supports ou technologies différents,
- dont 1 se trouve hors site.



Quels enjeux autour des données de la recherche ?



La science ouverte

La Science Ouverte et les données de la recherche:

- Développer une démarche proactive d'ouverture des données associées aux articles et publications
- Permettre la diffusion, l'accessibilité, l'interopérabilité et la pérennité des données de la recherche (FAIR)
- Garantir la reproductibilité des résultats



Le cadre national et international

Cadre national :

- Loi pour une **République Numérique de 2016**
- **Deuxième Plan National pour la Science Ouverte (2021-2024)** : ouverture des données de la recherche
- Politique Science ouverte de l'**ANR**
- **Printemps 2022** : lancement de l'entrepôt national des données de la recherche, **Recherche Data Gov**



Les entrepôts de données



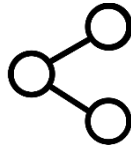
Un entrepôt de données permet de :



Stocker

Conserver sur un court et moyen terme 5 à 10 ans.

Garantir la pérennité des données de recherche.



Partager

Rendre **accessibles** vos données.

Attribuer un **identifiant pérenne** à vos données de recherche (DOI, PURL, ARK).



Diffuser

Faciliter la **réutilisation des données de recherche**.

Définir les **conditions d'accès** à vos données.



Comment choisir un entrepôt de données...





Trouver un entrepôt de données:



Voir Re3data : <https://www.re3data.org/>

Voir CatOpidor : <https://cat.opidor.fr/index.php/Conservation>

Critères pour choisir un entrepôt :

- Identifiant pérenne 
- Choix de licence juridique pertinent 
- Choix des modalités de diffusion des données
- Métadonnées adaptées
- Propriété des données conservée



Quelques entrepôts de données...





Le DAREC et ses services



Les données de la recherche

[Plan Science Ouverte](#) de l'UPHF : focus sur les publications, mais aussi incitation au dépôt des données en accès ouvert, et annonce de services à venir sur la gestion des données de la recherche

Diffusion d'un questionnaire auprès des chercheurs et laboratoires :

- Réaliser un état des lieux précis des usages, besoins et réalisations des chercheurs et laboratoires concernant les données de la recherche.
- Mener des entretiens ciblés.

Suite à la réalisation des entretiens et à l'analyse des résultats, le DAREC et l'UPHF vous proposeront des services adaptés à vos besoins.



Pour aller plus loin

Sur le cycle de vie de la donnée : <https://doranum.fr/enjeux-benefices/le-cycle-de-vie-des-donnees-de-recherche/>

Sur les principes FAIR : <https://doranum.fr/enjeux-benefices/principes-fair/>

Sur les métadonnées : <https://doranum.fr/plan-gestion-donnees-dmp/origine-description-donnees-recherche/>

Sur la question des standards de métadonnées :
<https://doranum.fr/metadonnees-standards-formats/standard-metadonnees/>



Pour aller plus loin

Sur le contexte national et européen de la science ouverte :

La politique Science Ouverte de l'ANR :

<https://anr.fr/fr/lanr/engagements/la-science-ouverte/>

Deuxième Plan national pour la science ouverte 2021-2024:

<https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>

La science ouverte dans les projets Horizon Europe, un critère d'excellence :

https://ec.europa.eu/info/funding-tenders/opportunities/docs/20212027/horizon/other/events/20210421/open-science_en.pptx



Pour tester vos connaissances

Jeu sur la gestion des données :

Valérie Gadrat, Yvette Lafosse, Claire Sowinski, & Coralie Wysoczynski. (2019, November 8). Jeu GopenDoRe : le jeu coopératif pour acquérir des bonnes pratiques de gestion et de partage des données de la recherche.

<https://doranum.fr/enjeux-benefices/gopendore-jeu-sur-la-gestion-des-donnees-de-recherche/>





Contactez-nous !

Pour toute question ou demande d'accompagnement, n'hésitez pas :

Contact :

Melissa.Defond@uphf.fr :

Responsable du Département d'Appui à la Recherche
Référente Science Ouverte pour l'UPHF

Eva.Goubert@uphf.fr :

Chargée de mission données de la recherche
Département d'Appui à la Recherche