

Atelier :

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

FCC – 06 décembre 2022
SCD Université Polytechnique Hauts-de-France

Qu'entend-on par données de la recherche ?	1
Quelques notions autour des données de la recherche	2
Apprendre à bien gérer ses données de recherche	3
Quels enjeux autour des données de la recherche ?	4
Présentation de Recherche Data Gouv	5
Le DAREC et ses services	6
Pour aller plus loin	7

Plan

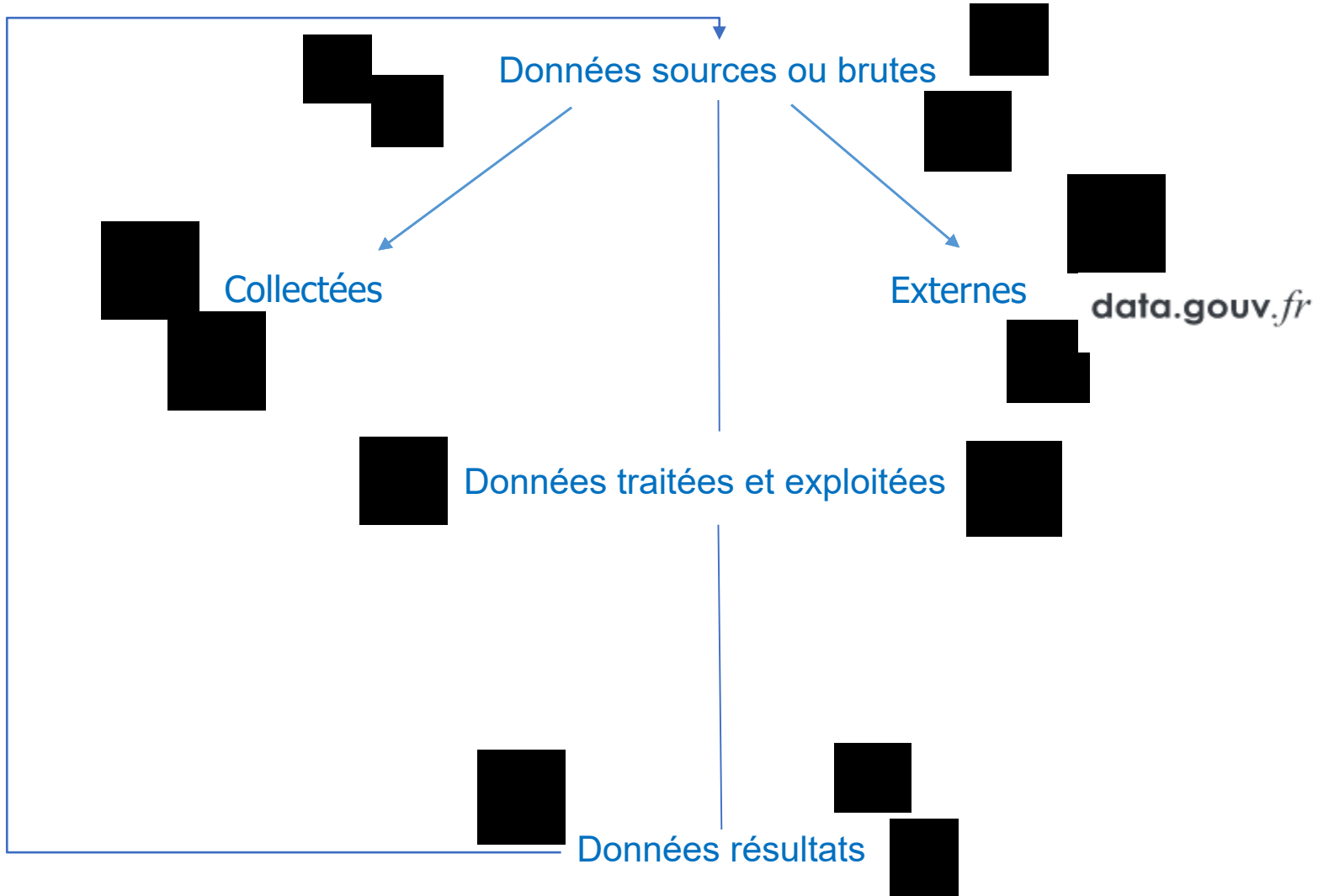
Qu'entend-on par données de la recherche ?



Qu'entend-on par 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* »



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

Définition des données de la recherche

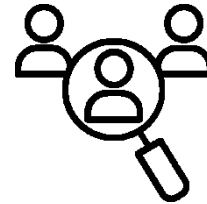
Données sources : Données qui constituent les matériaux de base de la recherche, avant tout traitement et analyse.

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

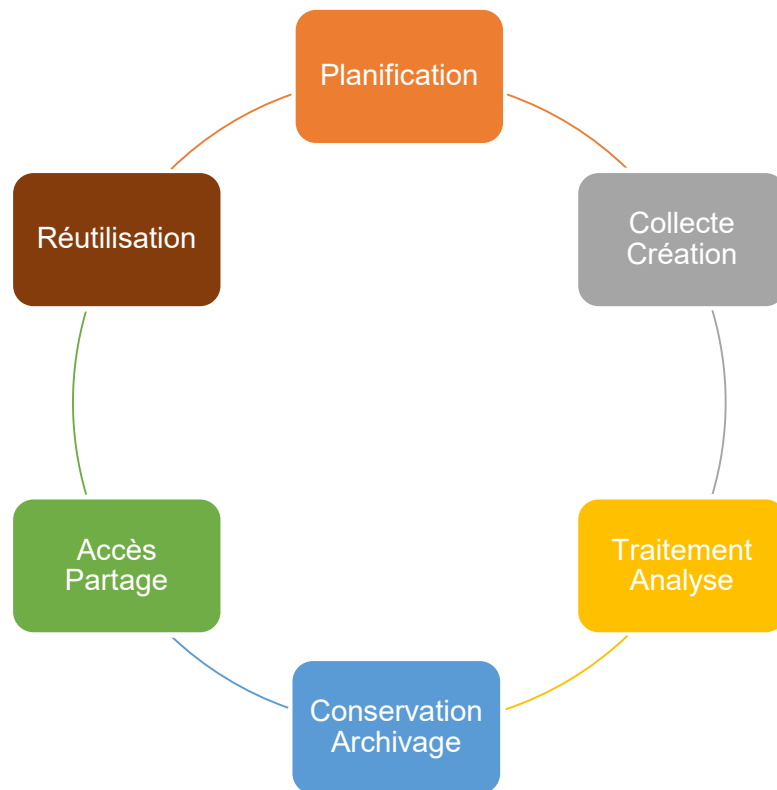


Quelques notions autour des données de la recherche...

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.



Quelques notions autour des données de la recherche...

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

Quelques notions autour des données de la recherche...

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

Quelques notions autour des données de la recherche...

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/>

Quelques notions autour des données de la recherche...

Le Plan de gestion de données

ANR : 6 rubriques

[Trame ANR](#)

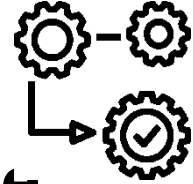
1. Description des données.
2. Documentation et qualité des données.
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données et conservation à long terme
6. Responsabilités et ressources

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 problèmes neurolocomoteurs. Elle a obtenu d'excellents résultats qu'elle 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.

Comment bien gérer ses données de recherche ?

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.

Outil : [Metada Standards Catalog](#)

Comment bien gérer ses données de recherche ?

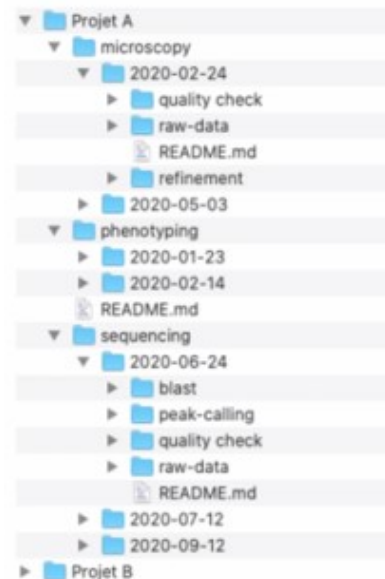
Focus : Organiser et nommer ses données

Mettre en place un plan de classement

Permet de faciliter le classement et le repérage des données et documents.

Affecter chaque donnée à un dossier précisément identifié.

- Réaliser une arborescence hiérarchisée.
- Partir du général vers le particulier.
- Refléter votre activité.



Comment bien gérer ses données de recherche ?

Focus : Règles de nommage des fichiers

Deux principaux objectifs :

- Garantir la pérennité des données
- Permettre la compréhension des informations

Comment nommer les fichiers ?

- Noms succincts, pas d'espace ou de caractères spéciaux.
- Utiliser des noms de fichiers précis et explicites.

Par exemple :

- Date
- Sujet
- Type de données
- Version

Comment bien gérer ses données de recherche ?

Documenter ses données (Tracer, Indexer)

Pourquoi ?

- Préparer le partage et l'interopérabilité.
- Optimiser le stockage et l'archivage.
- Faciliter la recherche des données de recherche et publications.

Comment ?

- En produisant de la documentation sur vos données (métadonnées, document, fiche synthétique...).
- Réaliser dès l'étape de collecte et de création des données de la documentation sur vos données de recherche.

Comment bien gérer ses données de recherche ?

Focus : Fichier Read me

- Associer aux données un fichier README
- Décrit les données pour les rendre compréhensibles et réutilisables.
- On y retrouve les instructions d'interprétation et de réutilisation des données.

Exemple concret ?

[Fichier Read Me Cornell University](#)

[Guidelines for creating a README file](#)

Comment bien gérer ses données de recherche ?

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 ?

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

Quels enjeux autour des données de la recherche ?

Le cadre national et international

Cadre européen :

- Agences européennes encouragent fortement le dépôt des données issues des travaux qu'elles financent.
- **Plan Horizon Europe qui succède à H2020.**



Quels enjeux autour des données de la recherche ?

Le cadre national et international

Cadre national :

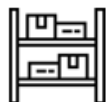
- Loi pour une **République Numérique de 2016**
- **Deuxième Plan National pour la Science Ouverte** (2021-2024) : ouverture des données de la recherche
- Politique Science ouverte de l'**ANR**
- **Juillet 2022** : lancement de l'entrepôt national des données de la recherche, [Recherche Data Gov](#)



Quels enjeux autour des données de la recherche ?

Les entrepôts de données

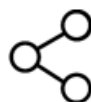
Un entrepôt de données permet de :



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.

Quels enjeux autour des données de la recherche ?

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

Quels enjeux autour des données de la recherche ?

Quelques entrepôts de données...



Quels enjeux autour des données de la recherche ?

L'entrepôt national des données de la recherche...



- Permet à toute personne de l'ESR de documenter et de partager tout type de données scientifiques.
- Facilite la recherche, la découverte, l'accès et la réutilisation des données de l'établissement dans le respect des bonnes pratiques et des règles éthiques et réglementation en vigueur.

Quels enjeux autour des données de la recherche ?

L'entrepôt national des données de la recherche...

Principaux cas d'usage :

Je monte un projet ANR avec des partenaires externes

Je publie un data paper

Je publie un article scientifique, l'éditeur me demande des données

Je recherche des données sur un sujet spécifique

Je veux rendre visibles des données

Je veux gérer et partager des données collectées et générées par ma structure

Je veux préserver et valoriser les données collectées par un collègue qui part de mon laboratoire

Je veux partager des données gérées par un autre système d'information



RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

recherche.data.gouv.fr

<https://entrepot.recherche.data.gouv.fr/>

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.

Quelques résultats...

85 réponses, **18%** de représentativité *source bilan social UPHF 2020*.



Avez-vous déjà rédigé un plan de gestion des données (PGD ou DMP) ?

58,82%

Non



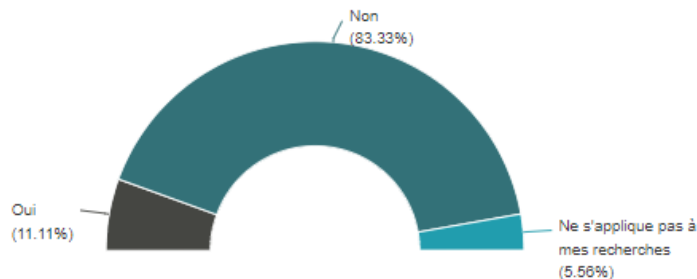
Connaissez-vous les principes FAIR (Findable, Accessible, Interoperable, Reusable) ?

84,71%

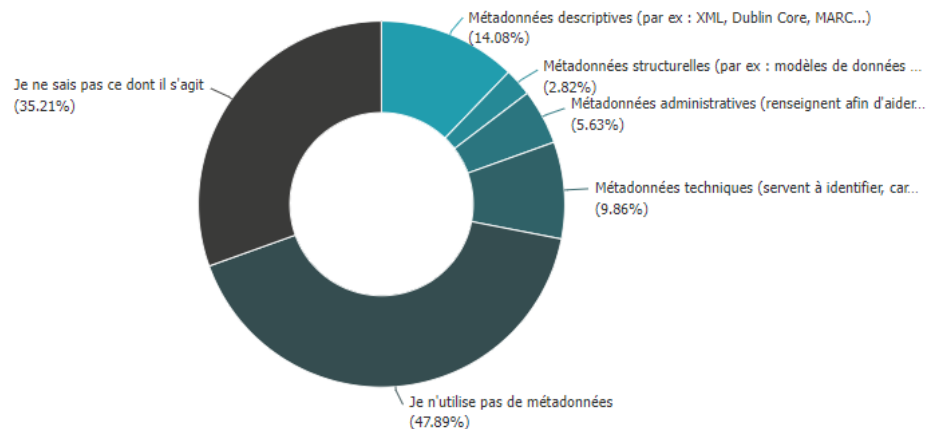
Non

Quelques résultats...

Connaissez-vous des directives ou recommandations pour la gestion des données de la recherche ?



La majorité des chercheurs n'utilisent pas de métadonnées.



Pour aller plus loin...

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

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

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

Sur la question des standards de métadonnées :
<https://dorum.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/2021_2027/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/>



Pour tester vos connaissances...

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 France Universités pour l'UPHF

Eva.Goubert@uphf.fr :

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

Merci pour votre attention

Merci pour votre attention