

Accompagner la vie des données

Caroline Doucoure, IST-IRD
caroline.doucoure@ird.fr



Plans de gestion des données

❖ Lâchez les freins

Accompagner le cycle de vie des données

❖ Ventilez les informations

Rédiger un DMP avec un outil en ligne

❖ TP



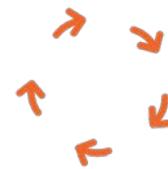
Accompagner le cycle de vie des données

- Se poser des questions à chaque étape**
- Pour bien commencer
- Suivre un modèle





Le cycle de vie de la donnée



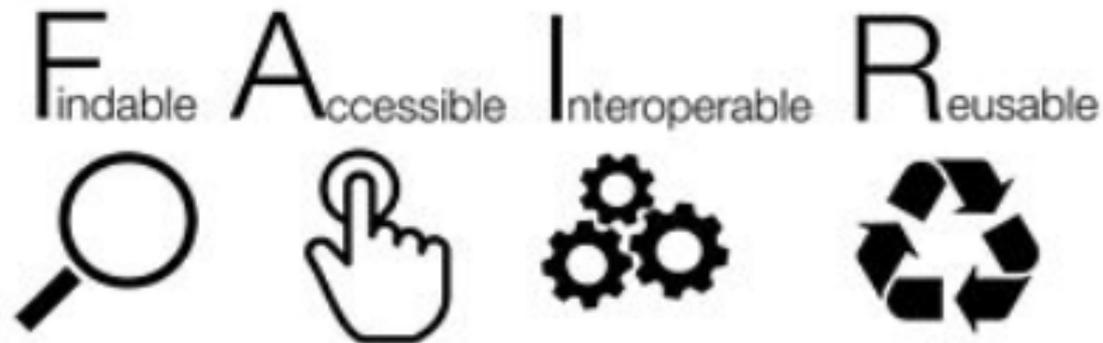
Le cycle de vie



Ouvert autant que possible



« **Aussi ouvert que possible, aussi fermé que nécessaire** »



[Dorandum : les principes FAIR : quels enjeux et bénéfices](#)



Pourquoi avoir choisi ce cycle de vie des données

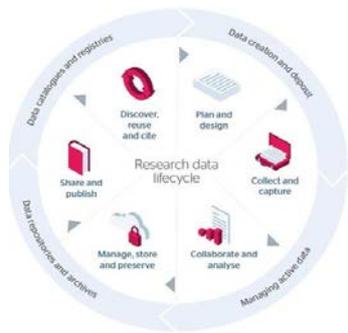
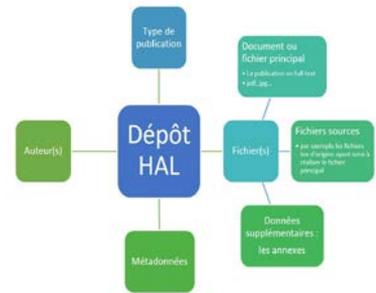
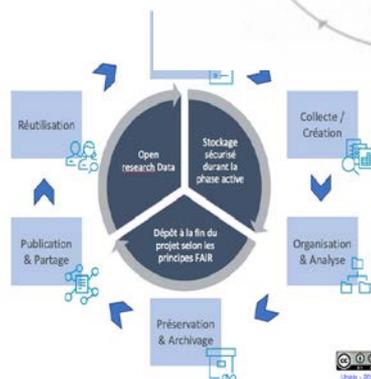
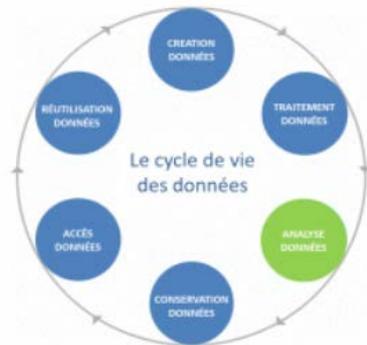


Schéma du cycle de vie des données
<https://data.ird.fr/utiliser-des-donnees/>



Pourquoi avoir choisi ce cycle de vie des données

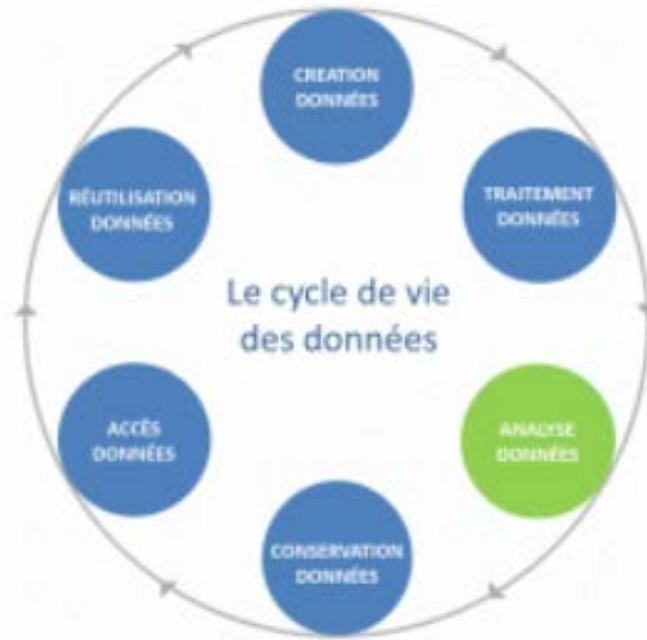


Schéma du cycle de vie des données de l'INIST CNRS

https://www.inist.fr/wp-content/uploads/donnees/co/module_Donnees_recherche_7.html

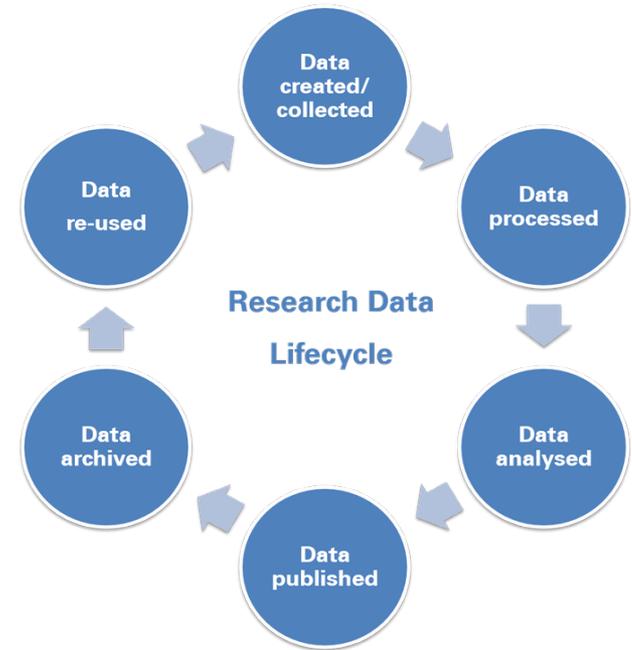


Schéma du cycle de vie des données (Université Dresden Allemagne)

https://tu-dresden.de/zih/dienste/service-katalog/zusammenarbeiten-und-forschen/forschungsdatenmanagement?set_language=en

Pourquoi avoir choisi ce cycle de vie des données

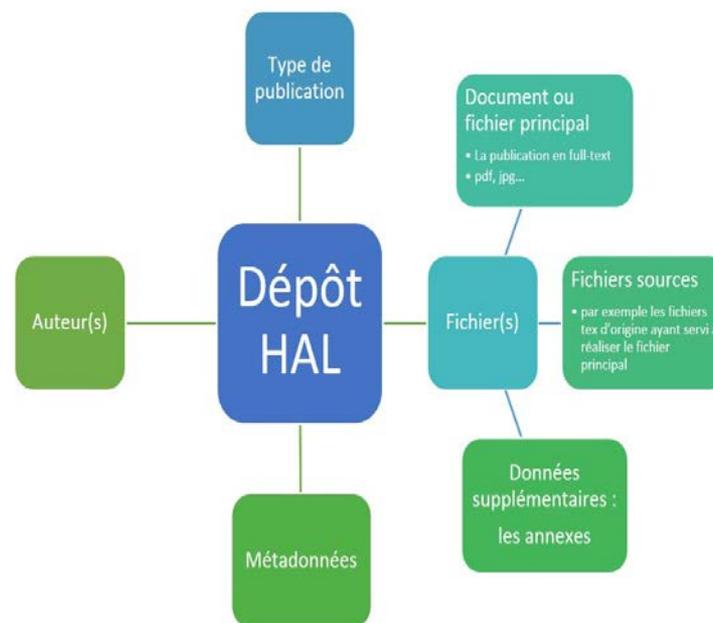


Schéma du cycle de vie des données d'Iqualit

<https://www.igualit.com/article/quel-est-le-cycle-de-vie-de-vos-donnees-electroniques/>

Schéma du cycle de vie des données du CCSD CNRS

https://memsic.ccsd.cnrs.fr/mem_01374509/

Pourquoi avoir choisi ce cycle de vie des données

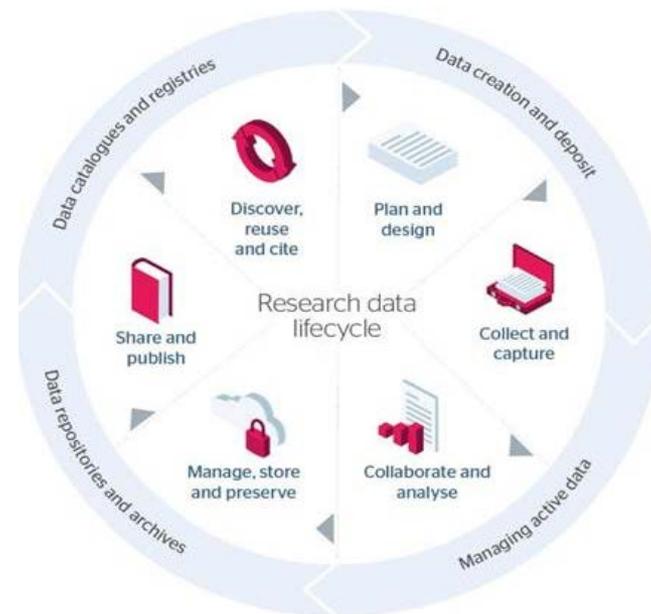
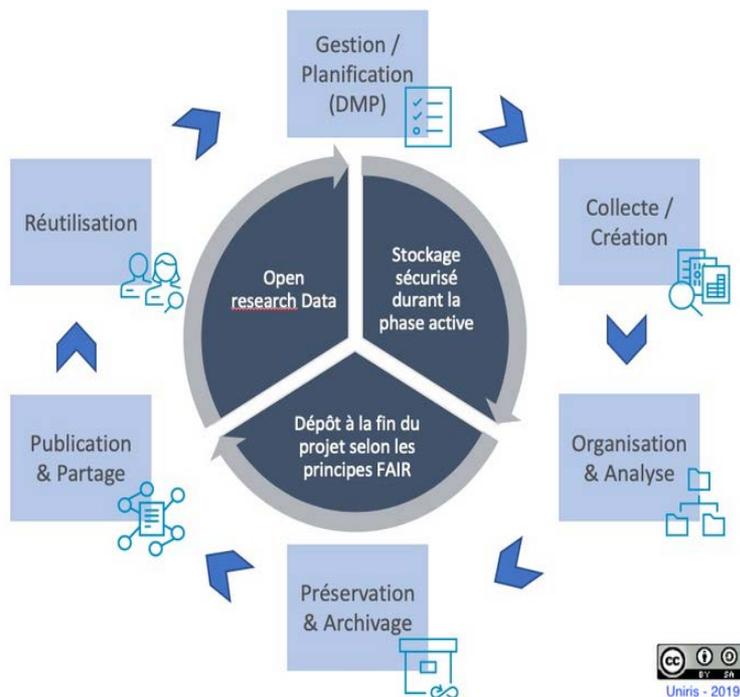


Schéma du cycle de vie des données (Université Lausanne Suisse)

<https://www.unil.ch/openscience/fr/home/menuinst/open-research-data/les-donnees-de-recherche/cycle-de-vie-et-types-de-donnees.html>

Schéma du cycle de vie des données de Researchgate

https://www.researchgate.net/figure/The-Research-Data-Lifecycle-https-rdmtoolkitjiscacuk-research-datalifecycle_fig1_339659891



Le cycle de vie des données

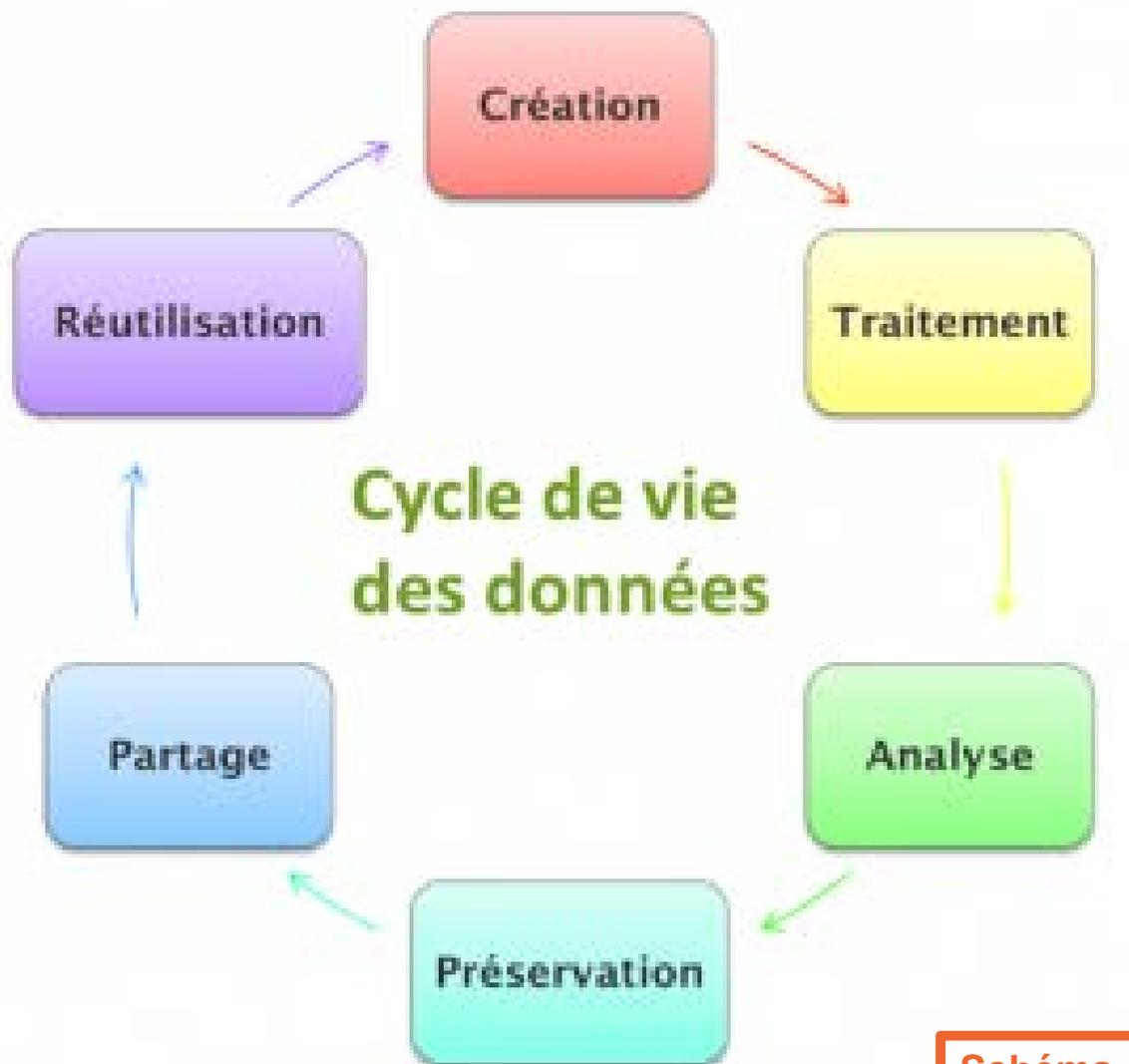


Schéma du cycle de vie des données
<https://data.ird.fr/utiliser-des-donnees/>



Le DMP et le cycle de vie des données



Schéma du cycle de vie des données

<https://www.ouvrirlascience.fr/passeport-pour-la-science/guide-pratique-a-lusage-des-doctorants/>



Le DMP : des questions à se poser

- Quels sont les besoins générés par la gestion des données ?
- Comment seront-ils financés, à court et à long terme ?

Ressources

- En quoi consiste le projet ?
- Qui sont les partenaires ?
- Quelle est la politique de gestion des données ?
- Qui est responsable de la gestion des données ?

Responsabilités dans le projet

- Quelles données seront produites/utilisées au cours du projet ? (type, format, volume et accroissement...).
- Comment seront-elles produites ou transformées ?

Collecte des données

- Comment, où, par qui, seront stockées, sauvegardées et sécurisées les données au cours du projet ?

Sauvegarde des données

- Qui pourra accéder aux données ? Quelles données seront partagées ? publiées ?
- Avec qui ?
- Comment ?
- Dans quel délai ?
- Sous quelle licence ?

Accès et partage des données



- Comment les données seront-elles identifiées, décrites ?
- Quels standards de métadonnées utilisera-t-on ?

Documentation des données

- Quel plan pour l'archivage et la préservation à long terme ?
- Pour quelles parties des données ?

Archivage et préservation des données

- Des données sensibles seront-elles produites ou utilisées ?
- Comment sera assurée leur anonymisation ?

Ethique

- Qui sera propriétaire des données produites ?
- Des données externes seront-elles utilisées ?

Propriété intellectuelle

**Les Plans de gestion de données
/S. Cocard et D. L'Hostis (INRA), URFIST, juillet 2018**

Accompagner le cycle de vie des données

- ✓ Se poser des questions à chaque étape
- Pour bien commencer**
- Suivre un modèle



Pour bien commencer



Suivre les conseils de sites de référence
([Doranum](#), [CoopIST](#), [IRD-Data](#))

Respecter les standards de sa communauté

Se renseigner : que demande le financeur (modèle, calendrier, grille d'évaluation ...)

Consulter des exemples de DMP :
https://dmp.opidor.fr/public_plans/





Accompagner le cycle de vie des données

- ✓ Se poser des questions à chaque étape
- ✓ Pour bien commencer
- Suivre un modèle**

Accompagner le cycle de vie des données

- ✓ Se poser des questions à chaque étape
- ✓ Pour bien commencer

☐ **Suivre un modèle**

- 1. Description des données : collecte ou réutilisation de données existantes**
2. Documentation et qualité des données
3. Stockage et sauvegarde pendant le projet
4. Exigences légales et éthiques, codes de conduite
5. Partage des données et conservation à long terme
6. Responsabilités et ressources



Définition : les données scientifiques (données de recherche, Research Data)



Pour l'OCDE (2006) :

« **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 la recherche. »



Ne sont PAS des données de recherche



Selon l'OCDE :

Les documents non achevés :

- les carnets de laboratoire
- les analyses préliminaires, les projets de documents scientifiques et les programmes de travaux futurs

Les objets matériels, les animaux et les matériaux vivants :

- Les échantillons
- les souches bactériennes et les animaux de laboratoire

Les productions scientifiques secondaires :

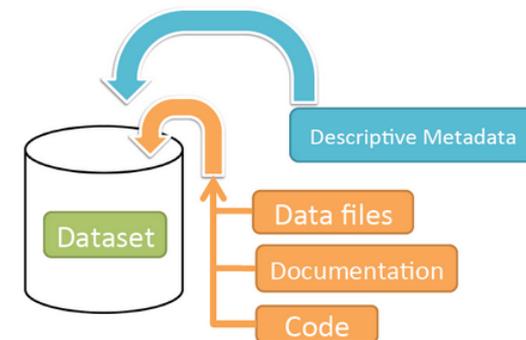
- les publications scientifiques, les communications à congrès
- les supports de formation

Les données administratives non intégrées dans un corpus de recherche...



Définition : Jeu de données (DataSet)

- « **Agrégation**, sous une forme lisible, de données brutes ou dérivées présentant **une certaine unité**, rassemblées pour former **un ensemble cohérent**. » (Gaillard, 2014).
- **Unité de dépôt dans un entrepôt** : sa granularité ne doit être ni trop petite, ni trop grande, pour faciliter la citation.
- **Complété par des métadonnées** nécessaires à leur réutilisation et à leur validation : méthodes d'obtention, logiciels nécessaires, auteurs, couverture temporelle et spatiale, etc.

Schematic Diagram of a **Dataset** in Dataverse 4.0

Container for your data, documentation, and code.

On peut agréger plusieurs **jeux de données** dans un même **produit de recherche**.



Présenter l'ensemble des jeux de données (conseil : sous forme de **tableau**) avec pour chaque jeu : nom de la source, format du fichier, identifiant, description du contenu, volume (un ordre de grandeur) ...

<https://data.ird.fr/gerer/quelles-donnees>

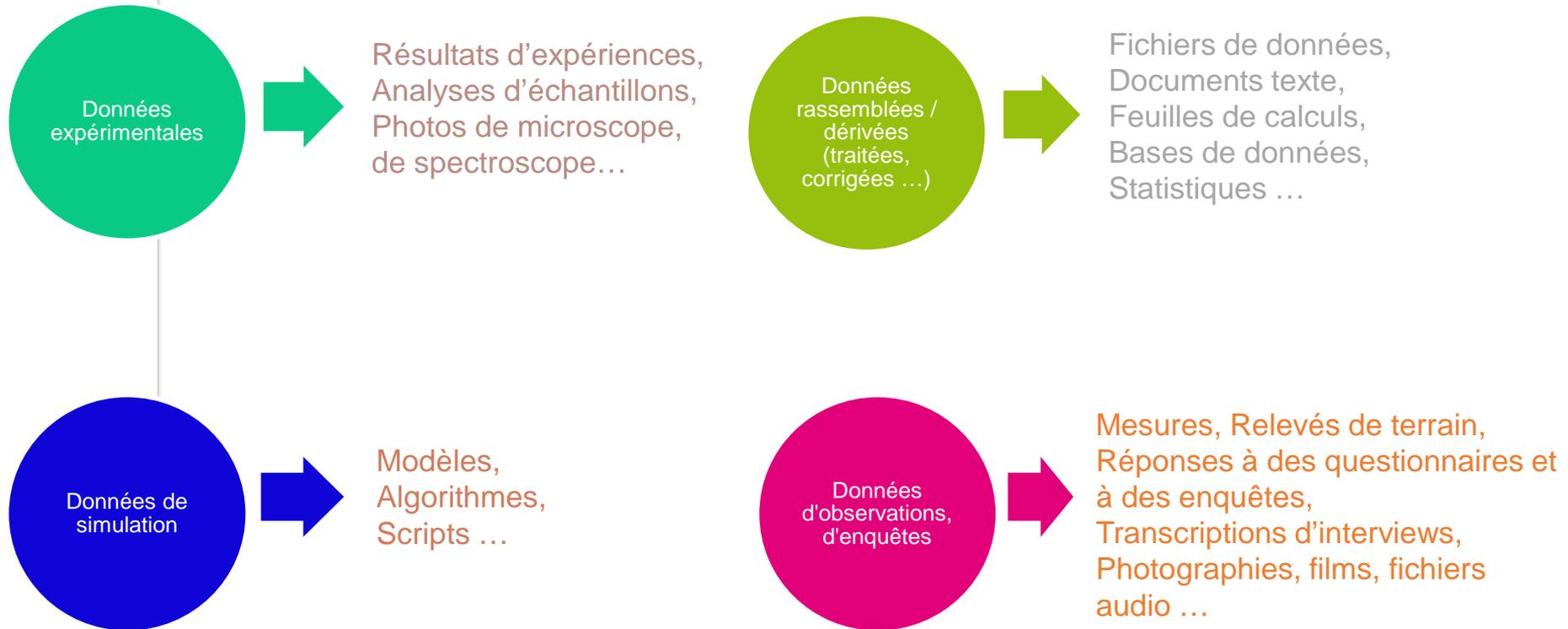
Formats de fichiers ouverts/fermés (Dorandum) : de préférence non-propriétaires, ouverts, pérennes et gratuits (réutilisation des données)

Types de fichiers (Datapartage INRAE)





Typologie / Méthodologie d'obtention



Accompagner le cycle de vie des données

- ✓ Se poser des questions à chaque étape
- ✓ Pour bien commencer

☐ **Suivre un modèle**

1. Description des données : collecte ou réutilisation de données existantes

2. Documentation et qualité des données

3. Stockage et sauvegarde pendant le projet

4. Exigences légales et éthiques, codes de conduite

5. Partage des données et conservation à long terme

6. Responsabilités et ressources





Métadonnées et documentation



Utiliser de préférence des standards de métadonnées généraux ou spécifiques à une discipline scientifique et des outils d'aide à la production de métadonnées :

<https://data.ird.fr/documenter-ses-donnees/>

Méthodologie de collecte

A minima, un fichier de type “lisez-moi” : nom de la source, format du fichier, identifiant, description du contenu ...





Qualité des données

Expliquer comment la qualité et la conformité de la collecte des données seront contrôlées et documentées :

<https://data.ird.fr/qualite-des-donnees/>

1. Présenter les objectifs à atteindre
2. Citer les processus qualité utilisés
3. Détailler les règles de **nommage et organisation** (Datapartage INRAE)
4. Préciser quels outils pratiques utilisés (checklists, cahiers de labo* ...)
5. Décrire le suivi régulier (hebdomadaire, mensuel ...) mis en place

* S'inspirer du guide des bonnes pratiques de Doranum pour tenir un cahier de labo : <https://www.dataacc.org/bonnes-pratiques/utiliser-un-cahier-de-laboratoire-numerique/ressource-doranum-cahiers-de-labo>



Accompagner le cycle de vie des données

- ✓ Se poser des questions à chaque étape
- ✓ Pour bien commencer

☐ **Suivre un modèle**

1. Description des données : collecte ou réutilisation de données existantes
2. Documentation et qualité des données
3. **Stockage et sauvegarde pendant le projet**
4. Exigences légales et éthiques, codes de conduite
5. Partage des données et conservation à long terme
6. Responsabilités et ressources



Stockage et sauvegarde pendant le projet



Les outils utilisés, de préférence des outils institutionnels, pérennes et régulièrement mis à jour

La procédure de sauvegarde au cours du projet : périodicité, automatisation, nombre de copies et localisation :

<https://data.ird.fr/sauvegarder-ses-donnees/>

Sauvegarde « 3.2.1 » : sauvegarde au moins 3 fois sur au moins 2 supports différents, dont 1 en ligne, de 3 versions successives

Les procédures d'accès aux données (mots de passe, numéros IP ...)



Accompagner le cycle de vie des données

- ✓ Se poser des questions à chaque étape
- ✓ Pour bien commencer

☐ **Suivre un modèle**

1. Description des données : collecte ou réutilisation de données existantes
2. Documentation et qualité des données
3. Stockage et sauvegarde pendant le projet
- 4. Exigences légales et éthiques, codes de conduite**
5. Partage des données et conservation à long terme
6. Responsabilités et ressources

Droit des données : le cadre juridique sénégalais



Le partage des données de recherche au Sénégal : les aspects éthiques et juridiques

Atelier juridique du lundi 25 octobre 2021

Une page sur le sujet a été créée sur le site OpenOPSE avec les liens vers les textes de lois, ou réglementaires, etc. s'appliquant au Sénégal :

➤ <https://lped.info/wikiObsSN/?AteJur>

- [Loi n°2008-09](#) portant loi sur le droit d'auteur et les droits voisins
- [Loi n°2012-03 du 03 janvier 2012](#) modifiant et complétant la Loi n° 2004-21 du 21 juillet 2004 portant organisation des activités statistiques

Droit des données : le cadre juridique sénégalais



Protection des données personnelles

[Loi n°2008-12](#) portant sur la Protection des données à caractère personnel

[Site de la CDP](#), Commission de Protection des Données Personnelles

[Formulaire de la CDP](#), demandes et déclarations





Droit des données : le cadre juridique français



- D'une manière générale, **les données sont réputées « de libre parcours »** : elles ne sont pas protégées par le droit d'auteur
- **Les données issues de la recherche sont considérées comme des documents administratifs**, si elles ont été produites dans le cadre d'une mission de service public ou majoritairement grâce à des fonds publics :
 - Elles sont donc **communicables à la demande** (sauf exceptions légales), si elles sont "**achevées**".
 - Elles sont soumises à un principe d'**ouverture par défaut**.
 - Elles doivent être diffusées **gratuitement et librement réutilisables**

Cf. [Loi pour une République Numérique \(2016\)](#)





Cadre juridique français

Exceptions – Publication sous condition

1. Les données **personnelles** ([RGPD](#)), en particulier les données **sensibles** ([CNIL](#)), dont les **données de santé** (secret médical professionnel) – cf [IRD Data](#)
2. Les données protégées par le [droit d'auteur](#) (œuvres originales)
3. Les données qui impliquent **un partenaire étranger ou privé** ([Droit sui generis des bases de données](#))
4. Les données concernant les **ressources génétiques et « connaissances traditionnelles » associées** : [Protocole de Nagoya](#)
5. Les informations pouvant avoir un impact sur la **conservation de la biodiversité** [Code de l'Environnement](#)

cf <https://data.ird.fr/cadre-juridique/>





Respect des aspects éthiques



Si certaines données peuvent soulever des problèmes d'éthique ou de déontologie, préciser quel comité sera contacté.

- Le Sénégal dispose d'un comité national d'éthique et de recherche en santé ([CNEERS](#)) et d'un comité d'éthique institutionnel à l'Institut de recherche en santé de surveillance épidémiologique et de formation ([IRESSEF](#)).
- L'IRD dispose d'un comité d'éthique pour la recherche en partenariat ([CCERP](#)) et est membre du comité consultatif commun d'éthique Inrae-Cirad-Ifremer-IRD .



Accompagner le cycle de vie des données

- ✓ Se poser des questions à chaque étape
- ✓ Pour bien commencer

☐ **Suivre un modèle**

1. Description des données : collecte ou réutilisation de données existantes
2. Documentation et qualité des données
3. Stockage et sauvegarde pendant le projet
4. Exigences légales et éthiques, codes de conduite
- 5. Partage des données et conservation à long terme**
- 6. Responsabilités et ressources**



Après le projet

Partage et conservation à long terme



Quelles données conserver ? (DCC)

Mes données sont-elles partageables ? (aide à la décision sur IRD Data)

Si on ne peut pas ouvrir les données, peut-on **partager les métadonnées** ?

Où publier/déposer mes données ?

Comment choisir un entrepôt ? (Re3data) cf. Formation J3

Choix d'une licence (Datapartage INRAE)

Après le projet

Partage et conservation à long terme



Une étude des entrepôts de données/datacenters/portails de données/géoportail/ entrepôts bibliographiques... existants dans les domaines d'OPSE sera menée en 2022, avec les critères de qualification des entrepôts.

En attendant, une page dédiée (liste d'entrepôts) a été créée sur le site **OpenOPSE** : <https://lped.info/wikiObsSN/?InvEnt>

Un webinaire sur le sujet pourra être prévu en parallèle avec la présentation de chacun des entrepôts par les intervenants correspondants.

Accompagner le cycle de vie des données

- ✓ Se poser des questions à chaque étape
- ✓ Pour bien commencer

☐ **Suivre un modèle**

1. Description des données : collecte ou réutilisation de données existantes
2. Documentation et qualité des données
3. Stockage et sauvegarde pendant le projet
4. Exigences légales et éthiques, codes de conduite
5. Partage des données et conservation à long terme
- 6. Responsabilités et ressources**

Rôles et responsabilités



Décrire les rôles et les responsabilités dans l'équipe projet concernant les activités de gestion des données : saisie des données, production des métadonnées, qualité des données, stockage et sauvegarde, archivage et partage des données ; mise en œuvre, examen et révision du DMP

Si possible nommer la(es) personne(s) responsable(s), prévoir des suppléants pour les rôles critiques

Préciser comment s'effectue la coordination des responsabilités et les modalités de décision entre partenaires



Budget et temps alloués



Budget : frais de stockage, de coût matériel, de temps de personnel, de coûts de préparation des données pour le dépôt, de frais d'entrepôt et d'archivage

Mentionner également les services obtenus gratuitement (par exemple [DataCirad](#) ou [DataSuds](#))

Ne pas oublier de chiffrer le salaire du/des gestionnaire(s) de données : <https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>

Le coût élevé de l'anonymisation des données est à anticiper le cas échéant



Conclusion



La minute Plan de gestion de données (2'57")



<https://doranum.fr/plan-gestion-donnees-dmp/minute/>



Merci pour votre attention

Q/R

A vous la parole ! Des questions ?