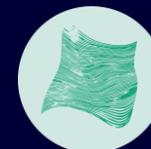


# Atelier Plan de Gestion de données *express*

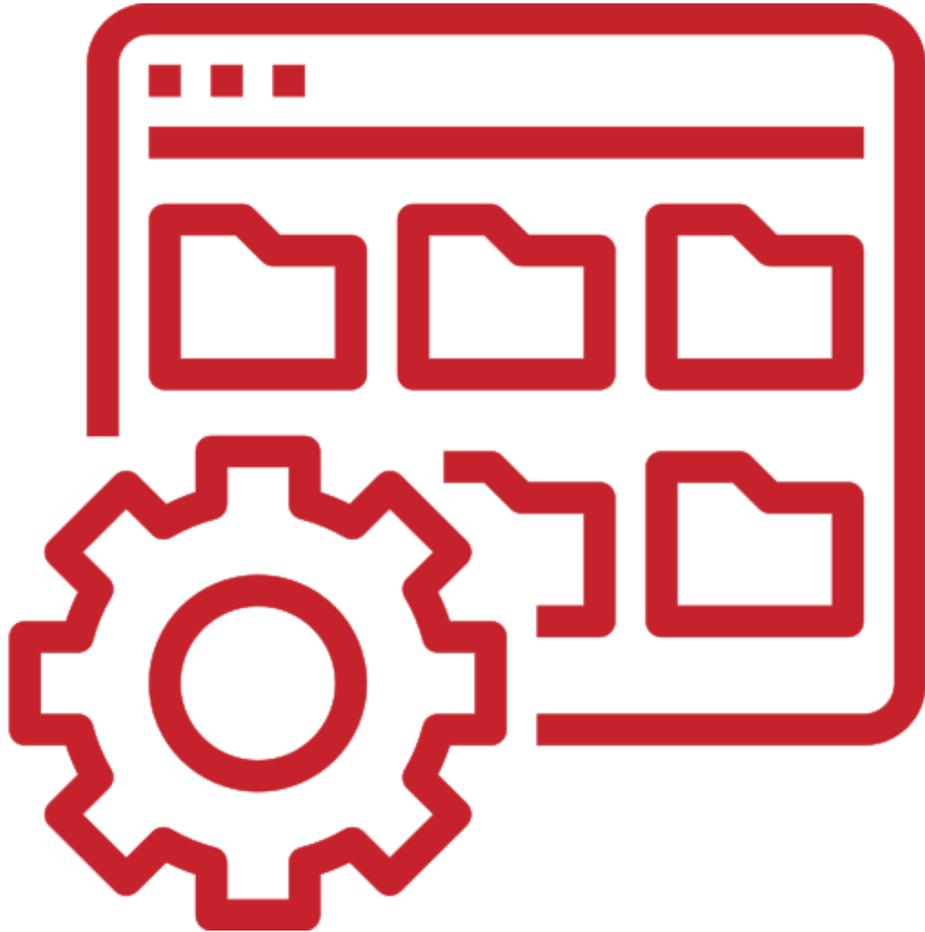
Mickael Malandran,  
Marie Cros

26 mai 2023



## Concernant les données : quelles sont les obligations ?

- 1) Rédiger un plan de gestion de données (PGD), ou *Data management plan* (DMP)
- 2) Avoir une gestion FAIR (*Findable, Accessible, Interoperable, Reusable*) de ses données
- 3) Ouvrir les données quand cela est possible (« aussi ouvert que possible, aussi fermé que nécessaire »)



# Les « données de recherche »

Enjeux, définitions

## Les données de recherche

- « **Enregistrements factuels** [...] utilisés comme **sources principales** pour la recherche scientifique et généralement reconnus par la communauté scientifique comme **nécessaires pour valider les résultats** de la recherche. »

(Principes et lignes directrices pour l'accès aux données de la recherche financées sur fonds publics OCDE, 2007)

- **Jeux de données (ou data set) :**

« Agrégation, sous une forme lisible, de données organisées pour former un ensemble cohérent »

Caractéristiques d'échantillons biologiques



Code source logiciel

```

<div class="container">
  <div class="row">
    <div class="col-md-6 col-lg-8">
      <div id="nav" role="navigation">
        <ul>
          <li><a href="index.html">Home</a></li>
          <li><a href="home-events.html">Home Events</a></li>
          <li><a href="multi-col-menu.html">Multiple Column Menu</a></li>
          <li class="has-children">
            <ul>
              <li><a href="tall-button-header.html">Tall Button Header</a></li>
              <li><a href="image-logo.html">Image Logo</a></li>
              <li class="active"><a href="tall-logo.html">Tall Logo</a></li>
            </ul>
          </li>
          <li class="has-children">
            <ul>
              <li><a href="variable-width-slider.html">Variable Width Slider</a></li>
              <li><a href="testimonials.html">Testimonials</a></li>
            </ul>
          </li>
        </ul>
      </div>
    </div>
  </div>
</div>

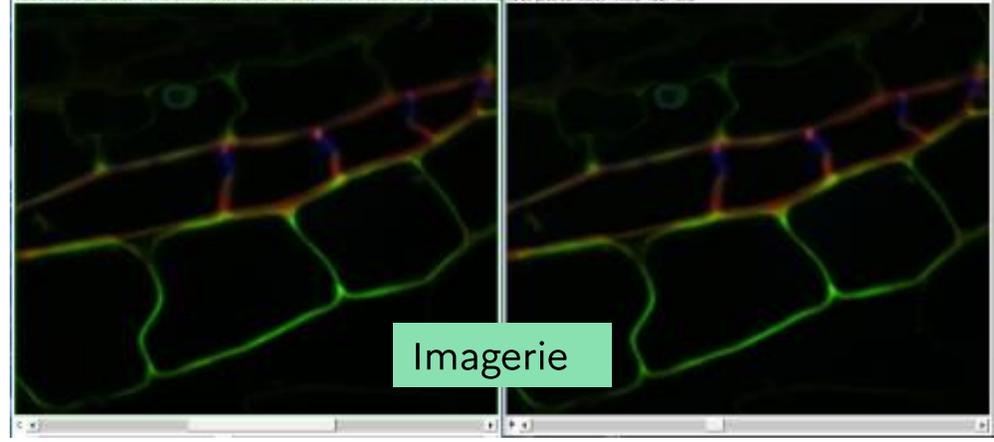
```

Actif occupé	Travail occasionnel	Durée du travail identique enq. précéd.	Temps de travail dans l'emploi principal	Raison principale travail temps partiel	Type d'horaires de travail	Nombre moyen de jours travaillés par semaine	Travail à domicile	Travail le soir (20H à minuit)	Travail de nuit (minuit à 5H)	Travail le samedi
1										
2	1	1			1	5.00	3	3	3	3
1										
1										
2	1	1			1	5.00	3	3	3	3
1										
1										
2		1						3	3	3
2		1			2	7.00	3	2	2	2
1										
2	2	1			1	6.00	2	2	3	1

Bases de données d'enquêtes

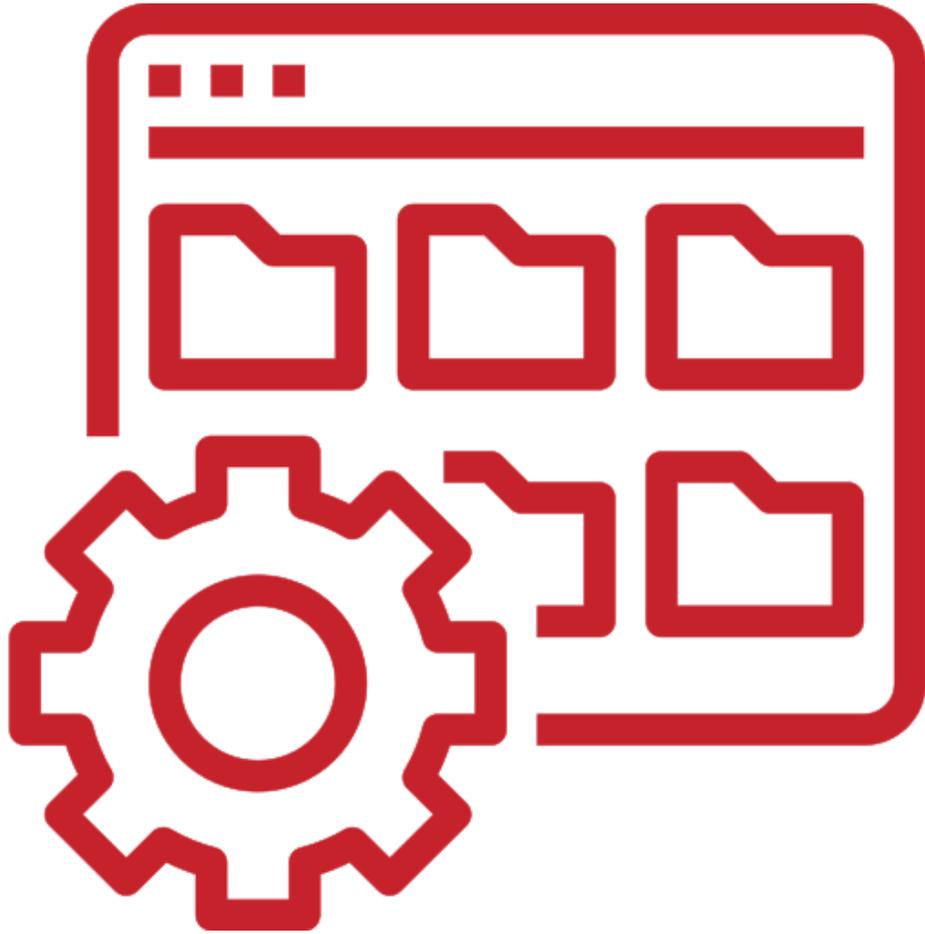
... la benda ritmo. Serà  
 ... dia bendà (fòto) que los comedians  
 ... legirà per far passar los dialògs  
 ... dins la lenga d'arribada.  
**Enregistrament** : Los comedians  
 enregistran los novèls dialògs jos la  
 direccion del cap de platèu.  
**Mesclatge** : Los dialògs son inter-  
 nacionals dins çò que se ditz la Version  
 internacionala (VI), que compren tots  
 s brutatges, la musica, los efèits,  
 ... Lo son deu èstre mesclat dins la  
 pècte estricte de la version  
 ... coma se lo film èra totjorn  
 ... dins la lenga d'arribada.  
**Back** : La novèla...

Corpus d'archives, textes, entretiens...



Imagerie

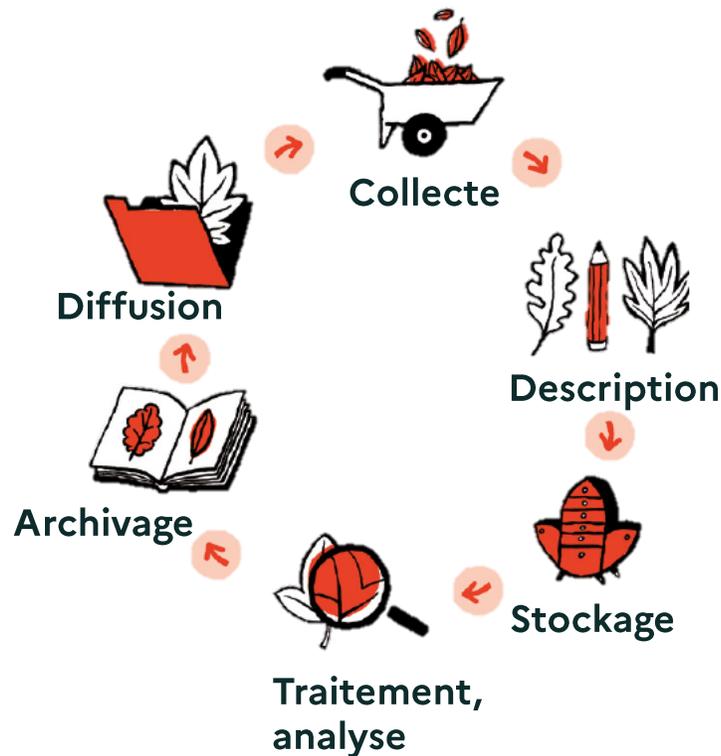
Mais aussi : résultats d'expériences et de mesures, données de carottage, séquences génétiques, données sismographiques, enregistrements audio ou vidéo, photographies, fichiers de paramètres machine...



## « Gérer » ses données

# Gestion des données

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



- Décrire les données collectées/crées
- Conserver les informations nécessaires à la compréhension
- Sauvegarder et stocker de manière sécurisée
- Partager
- Archiver
- Diffuser pour réutilisation

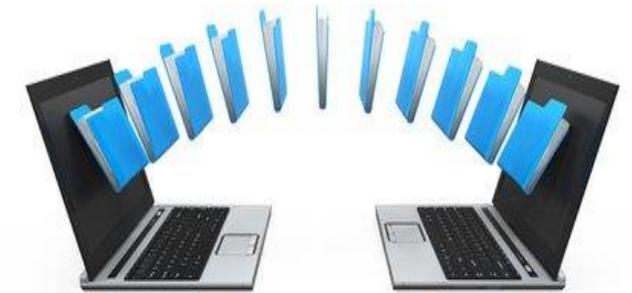
# Pourquoi « gérer » ses données ?



Gagner du temps



Limiter les risques



Partager / diffuser  
facilement

# « Bien » gérer ses données ?

Des données « FAIR » pour faciliter la découverte et la réutilisation

- Findable
- Accessible
- Interoperable
- Reusable



# Conseils et ressources



## La politique science ouverte de l'ANR

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

## FAQ de l'ANR sur le Plan de gestion de données

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

## La science ouverte dans les projets Horizon

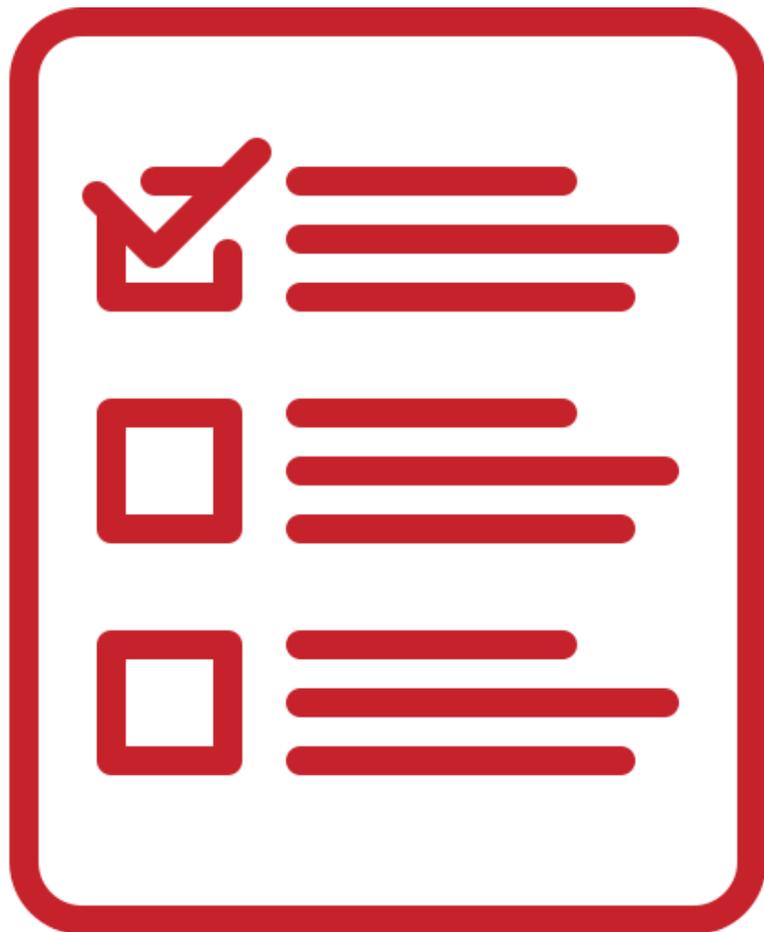
**Europe** [https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/other/events/20210421/open-science\\_en.pptx](https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/other/events/20210421/open-science_en.pptx)

## Les principes FAIR

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

## Gérer ses données de recherche, site web de l'université

<https://bu.univ-lille.fr/chercheurs-doctorants/science-ouverte/donnees-de-recherche>



# Le plan de gestion de données

*Data Management Plan*

## Le plan de gestion de données (PGD)

- **Document** qui décrit dès le début de la recherche comment seront gérées les données pendant et après le projet
- Il s'intéresse à l'ensemble du **cycle de vie des données** : collecte, traitement, analyse, partage, conservation, archivage, réutilisation
- Il sert à **formaliser** et à **partager ses pratiques** sur la gestion des données durant l'ensemble de leur cycle de vie
- Document **partagé** avec l'ensemble des parties prenantes du projet de recherche
- Livrable attendu par l'ANR et la commission européenne 6 mois après le début du projet, à mettre à jour au cours du projet.



# Contenu du plan de gestion de données

## ANR : 6 rubriques Trame ANR

- 1 - Description des données, collecte ou réutilisation de données existantes
- 2 - Documentation et qualité des données
- 3 - Stockage et sauvegarde pendant le processus de recherche
- 4 - Exigences légales et éthiques, codes de conduite
- 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

# DMP Opidor : une plateforme d'aide à la rédaction des PGD

- Se connecter : <https://dmp.opidor.fr/>
- Modèles de DMP
  - ANR ou Horizon Europe
  - Bientôt : U-Lille
- Créer un plan
- Consulter des plans de gestion de données rédigés
  - Sur la plateforme OPIDOR : « DMP publics »
  - [Catalogue de DMP de projets européens](#)




 Tableau de bord | Créer/Importer ▾ | DMP publics | **Modèles de DMP** | Aide | Plus ▾


 Université de Lille |  Université de Lille |  Nous contacter

Le Service Commun de Documentation de l'Université de Lille vous accompagne dans la rédaction de votre plan de gestion des données. [plus d'informations](#)

## Modèles de DMP

Modèles de DMP proposés par les financeurs ou par les organismes de recherche, disponibles dans DMP OPIDoR. Vous pouvez télécharger ces modèles et les recommandations associées, créer un plan à partir de ces modèles.

 Rechercher

Nom du modèle	Nom de l'organisme	Type d'organisme	Description	Dernière mise à jour	Télécharger	Créer un plan
ANR - DMP template (english)	Agence nationale de la recherche (ANR)	Financier	In line with its Open Science policy and the National Plan for Open Science, the French National Research Agency (ANR) requires all projects funded in 2019 onwards to produce a Data Management Plan (DMP). This move is intended to	16/05/2022	 	 





1

# Description des données : collecte ou réutilisation

## Description des données



- La partie description des données est structurante dans le PGD
- Elle va vous aider à identifier les principaux enjeux auxquels vous allez être confrontés : **stockage, accès, éthique, etc.**
- Il s'agit de décrire le plus précisément possible les données pour qu'elles soient compréhensibles et utilisables par les membres du projet, voire par d'autres chercheurs si vous partagez vos données à l'issue du projet.

# Description des données

*Les questions à se poser pour alimenter le PGD*

- **Est-ce qu'il s'agit de données...**
  - Produites ? Collectées ? Réutilisées ?
- **Quelles sont leurs caractéristiques ?**
  - **Types** : audio, vidéo, textes, tableaux, code, imagerie...
  - **Source/provenance** : comment ont-elles été produites ?
  - **Formats** : Quels formats sont utilisés ?  
privilégier les formats ouverts, pérennes et non propriétaires (jpeg, odt, tiff, pdf)
  - **Volume** : nombre d'objets, volume en octets
  - **Stabilité** : jeu de données achevé ? Croissant ?
  - **Valeur** : coût financier, en temps, en moyens
  - **Propriété** : données originales, données réutilisées
  - **Sensibilité** : données confidentielles, sensibles, médicales...



# Conseils et ressources



## Conseils pour la rédaction :

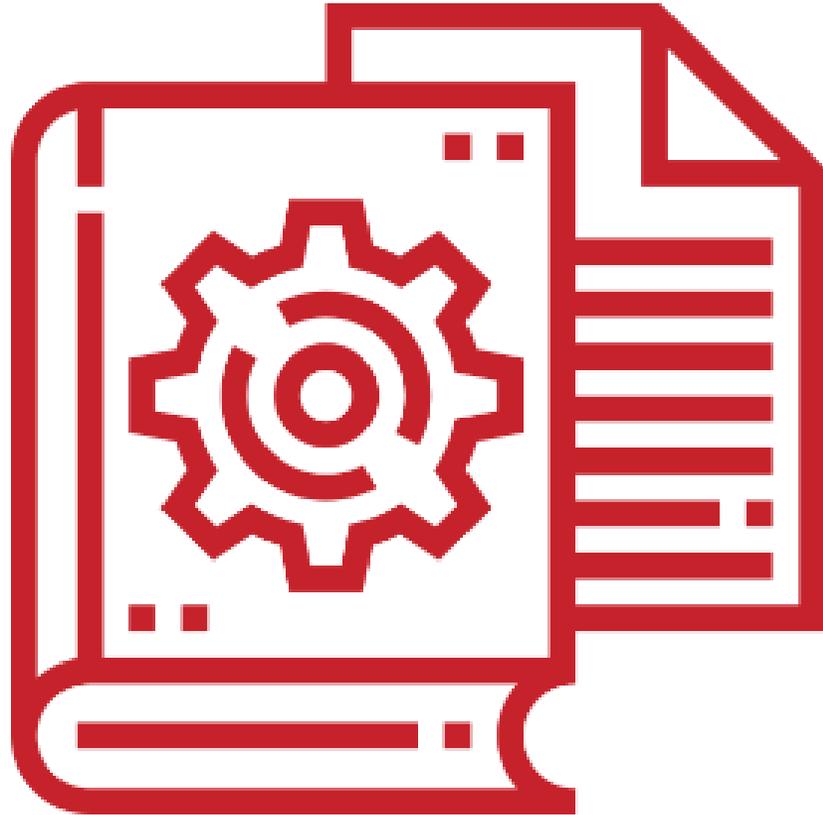
Identifier des sous-ensembles de données pertinents selon...

- ❖ Le type de données
- ❖ La provenance
- ❖ La propriété
- ❖ La stabilité
- ❖ ...

Dans chaque rubrique du PGD, décliner les réponses en reproduisant ce classement

## Des ressources pour aller plus loin :

- [Fiche Doranum](#) sur la description des données
- [Liste indicative de formats ouverts et fermés](#)
- [L'outil FACILE du CINES](#), pour vérifier si les formats de vos fichiers sont pérennes



2

## Documentation et qualité des données

## Documentation et métadonnées



- **Métadonnées :**

- Des données décrivant d'autres données
- Elles facilitent la recherche des données
- Elles aident à la compréhension des données et de leurs conditions de production

- **Documentation :**

- Ensemble des documents à diffuser avec les données, permettant d'en comprendre le **contenu** et le **contexte de production** pour en faciliter la **compréhension** et la **réutilisation**.

# Documentation et métadonnées

*Les questions à se poser  
pour alimenter le PGD*

- **Quelles sont les conditions de production ou de collecte des données ?**
  - Méthodologie, protocoles expérimentaux, traitements appliqués, logiciels ou instruments utilisés
- **Des standards de métadonnées sont-ils utilisés ?**
  - Data Documentation Initiative, Data Cite Metadata Schema, Dublin Core...
- **Des vocabulaires contrôlés sont-ils utilisés pour indexer et décrire les données ?**
  - Référentiels, thésaurus, URI...
- **Où les métadonnées se trouvent-elles ?**
  - Métadonnées embarquées (ex : imagerie)
  - Métadonnées externes : dictionnaire de codes, cahier de laboratoire, fichier « readme »
- **Comment les métadonnées sont-elles organisées ?**
  - Organisation des dossiers
  - Convention de nommage des fichiers
- **Comment la qualité des données est-elle assurée ?**
  - Contrôle de la cohérence des données, taille ou répétition des échantillons, calibrage des instruments...

# Conseils et ressources



## Conseils :

- Produire la documentation au fur et à mesure
- Le cas échéant, discuter des pratiques collectives concernant la documentation : nommage des fichiers, organisation des dossiers partagés, etc.
- Se renseigner sur les pratiques disciplinaires, les schémas de métadonnées utilisés

## Des ressources pour aller plus loin :

### Métadonnées :

- [Fiche Doranum](#) sur les métadonnées
- [Des schémas de métadonnées](#) (video)
- [Standards de métadonnées](#)

### Read me :

- [Guide de Cornell](#)

### Nommage :

- [Règles de nommage de documents](#)



3

## Exigences légales et éthiques



# Des réglementations spécifiques peuvent s'appliquer aux données



**Concernent tous les traitements** de données : réutilisation, collecte, conservation, diffusion...

Loi de 2012 relative aux recherches impliquant la personne humaine

Remplacer l'utilisation d'animaux par d'autres méthodes expérimentales  
**Réduire** le nombre d'animaux utilisés pour les expériences  
**Raffiner** (améliorer) les procédures expérimentales pour réduire la souffrance animale

- **Droit des bases de données**
- **Loi Jardé**
- **Expérimentations animales : les 3R**
- **Protocole de Nagoya**
- Code de la santé publique
- Secret statistique
- Code de la propriété intellectuelle
- Protection des données personnelles (RGPD)

**Droit des bases de données / Loi Lemaire**  
Notion de donnée publique => on ne peut faire obstacle à la réutilisation du contenu des bases de données que les administrations publient.

Accord international sur la **biodiversité**, portant sur l'utilisation des ressources génétiques de la planète

# Interlocuteurs spécialisés

## Délégué à la protection des données (DPO)

**Jean-Luc Tessier pour ULille**

Point de contact avec la CNIL. Personne en charge de la protection des données personnelles au sein de l'université de Lille, rôle d'expertise et de conseil.

[En savoir plus](#)

## Comité d'éthique pour la recherche (CER)

formule des avis et des recommandations pour mieux intégrer les questionnements éthiques aux projets de recherche. Propose une évaluation concrète de la conformité éthique des protocoles de recherche.

[En savoir plus](#)

## Comité de protection des personnes (CPP)

s'assure que tout projet de recherche biomédicale sur l'être humain mené en France respecte diverses mesures (médicales, éthiques et juridiques) visant à assurer la protection des personnes qui participent à cette recherche.

[En savoir plus](#)

# Ressources



## Intranet U-Lille

<https://intranet.univ-lille.fr/environnement-de-travail/donnees-personnelles/>

Les guides du délégué à la protection des données (disponibles sur l'intranet) :

- [Chiffrer son disque dur](#)
- [Directeur de thèse](#)
- [Réaliser une enquête anonyme](#)

## CNRS

[Pratiquer une recherche intègre et responsable](#)

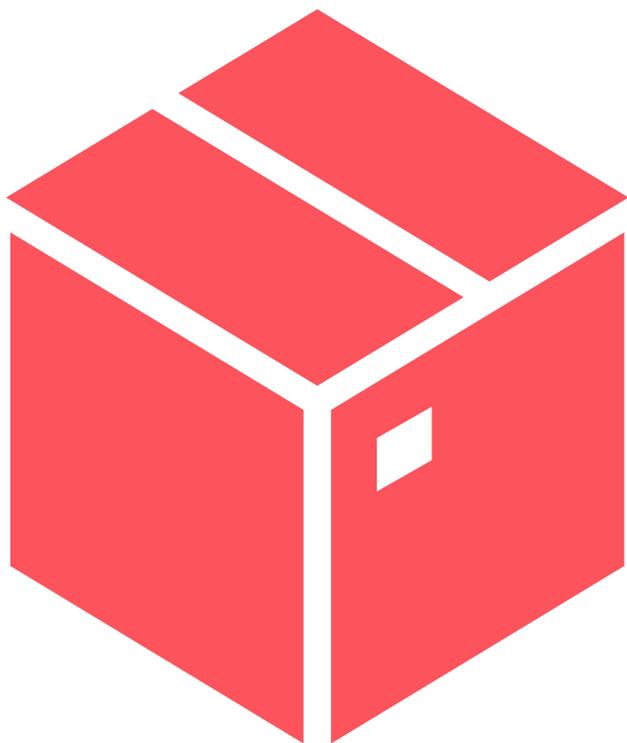
Comité d'éthique du CNRS (COMETS), 2017

**Doranum** « Aspects juridiques, éthiques, intégrité scientifique »

<https://doranum.fr/aspects-juridiques-ethiques/>

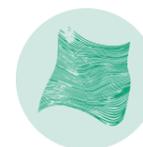
**Logigramme** – Questions juridiques liées à la diffusion des données (Institut Pasteur)

<https://www.pasteur.fr/fr/file/20707/download>

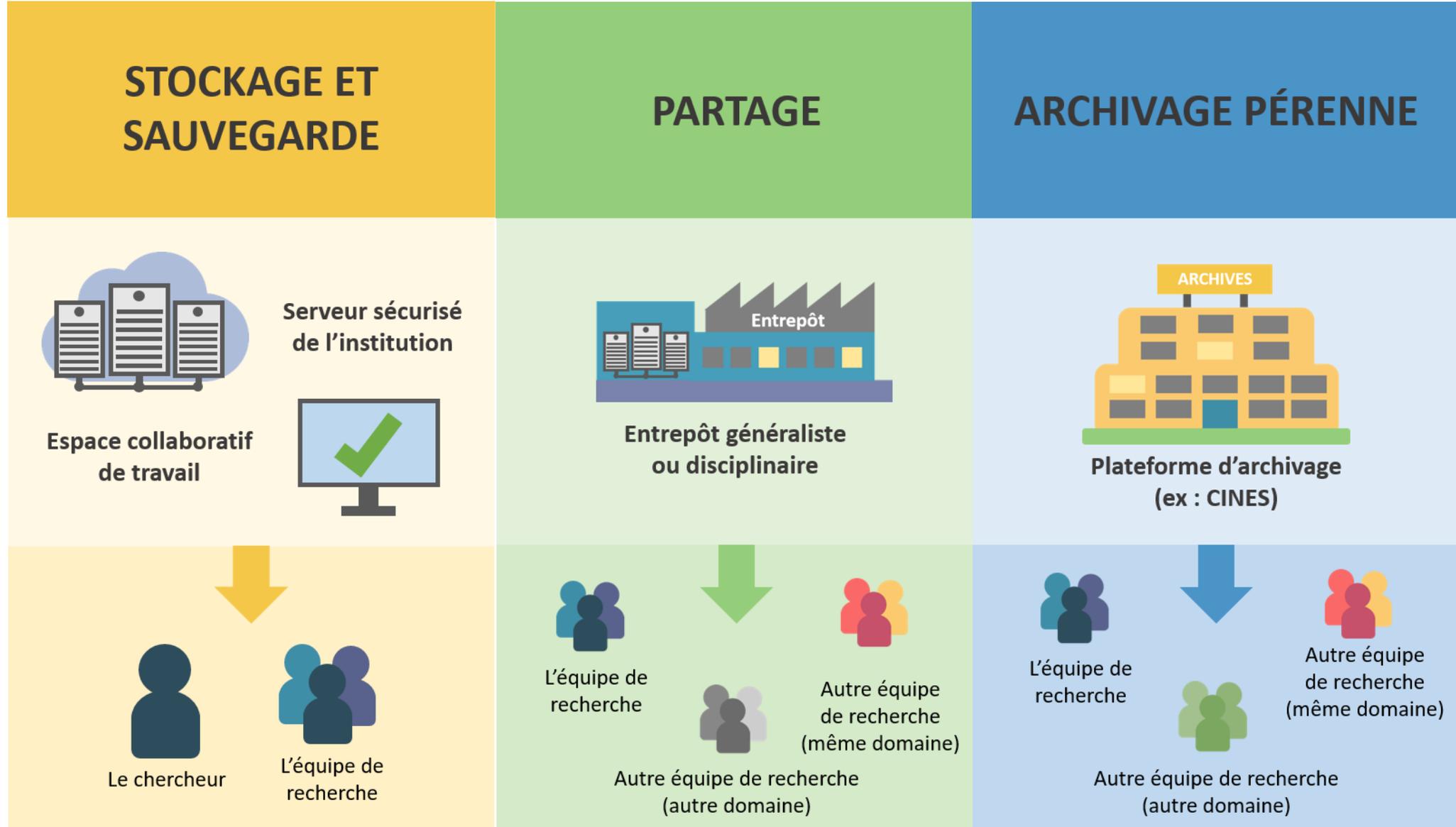


4

## Stockage et sauvegarde des données



# Vocabulaire : quelques précisions



# Le stockage des données

- La partie « stockage » du PGD vise à répondre aux enjeux liés à la sauvegarde et au partage au sein de l'équipe **pendant la durée de vie du projet.**



- En décrivant les méthodes utilisées, on prévient les **risques** de perte des données.

# Les enjeux du stockage et de la sauvegarde

- **Sécurité des données**
  - Ne pas perdre de données en cours de projet et pouvoir les récupérer en cas d'incident
  - Se fixer des règles pour limiter les risques
- **Protection des données**
  - contrôle des accès aux données (surtout pour des données sensibles)
  - politique et interlocuteurs pour protéger les données
- **Gagner du temps et de l'efficacité lors d'un travail collaboratif**
  - Transmettre des fichiers lourds
    - Outil : <https://filesender.renater.fr/>
  - Permettre différents niveaux d'usage d'un fichier (lecture simple, écriture, suppression)
  - Donner accès de manière adaptée (limitation dans le temps, mot de passe...)
- **Sobriété numérique, impact environnemental des données**
  - Réduire le volume unitaire des données pour limiter le stockage et les flux d'échanges



# Sauvegarder ses données

*Les questions à se poser pour alimenter le PGD*

- **Quel lieux de stockage utiliser ?**
  - répartition des données entre partenaires, supports physiques utilisés...
- **Quelle politique de sauvegarde ?**
  - processus utilisé, fréquence des sauvegardes, contrôle de sécurité...
- **Quelques conseils :**
  - Principe du **3, 2, 1**
    - Sauvegarder ses données sur **3** supports
    - dans au moins **2** lieux de stockage différents
    - avec **1** est une sauvegarde en ligne avec synchronisation
  - Fréquence de sauvegarde à adapter en fonction de l'activité



## Stocker vos données de manière institutionnelle

### Offre de la Direction du numérique Université de Lille

- NextCloud : stockage en ligne, synchronisation, dossiers et documents partagés - jusqu'à 50 Go, serveur en ligne
  - <https://infotuto.univ-lille.fr/fiche/nextcloud>
- Hébergement de serveurs virtuels
  - <https://infotuto.univ-lille.fr/fiche/hebergement-serveurs-virtuels>
- Location d'espace de stockage de grande capacité
  - 0,5 – 10 To : <https://infotuto.univ-lille.fr/fiche/location-despaces-de-stockage-de-grande-capacite>
- Stockage HPC (Calcul) :
  - adossé aux infrastructures de calcul intensif (mésocentre)  
> 200 To
  - <http://hpc.univ-lille.fr/>      [hpc@univ-lille.fr](mailto:hpc@univ-lille.fr)

Voir également l'offre existante au sein de vos laboratoires



**5**

# **Partage des données et conservation à long terme**



# Pourquoi diffuser ses données ?

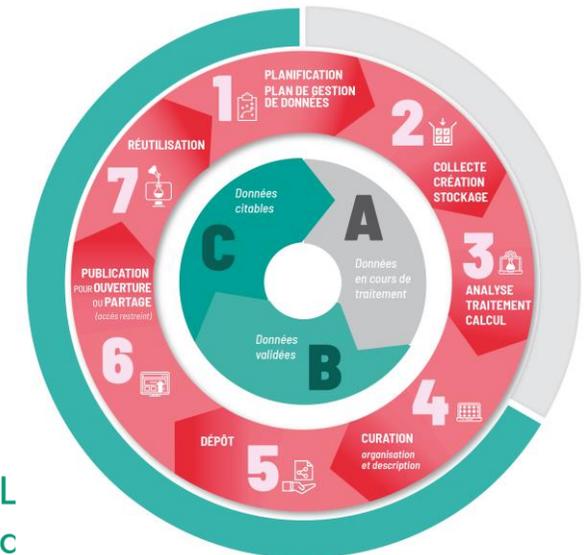
Le partage des données permet à d'autres équipes d'exploiter des jeux de données qu'elles n'auraient pas pu constituer en raison du **coût** de la collecte ou de la **méthode utilisée**.

Diffuser les données dans un entrepôt permet de s'appuyer sur une **structure compétente** pour les questions de :

- **partage** et de **conservation** des jeux de données
- **cadrer juridiquement** la réutilisation des données (contrats de confidentialité et licences d'utilisation)
- offrir une **plus grande visibilité** aux données partagées (attribution de DOI, moissonnage, etc.).



La loi pour une République Numérique de 2016 prévoit que si rien ne s'y oppose, les données produites grâce à des fonds publics (à hauteur de 50%) soient mises à disposition afin de sortir d'une logique de propriété privée et de constituer ces données en bien commun pour la communauté scientifique



# Diffuser en accès ouvert ou restreint ? Quelques règles

Le principe d'ouverture des données repose sur la maxime « *aussi ouvertes que possible, aussi fermées que nécessaire.* »

**À ce titre, restreindre l'accès n'est possible qu'avec les motifs suivants :**

Enjeux de confidentialité pour les personnes interrogées (données à caractère personnel)

Impossibilité d'anonymisation des données personnelles

Nécessité de produire d'abord des résultats d'analyses primaires

L'existence d'une propriété privée s'exerçant sur les données (parce qu'elles sont collectées en partenariat avec une structure privée)

Secret-défense

Secret industriel

Brevet

Embargo pour une valorisation scientifique des publications

# Licences d'utilisation

- Licences attribuées aux données, pour ses propres données et pour exploiter celles des autres
- Différence entre accès et réutilisation : accéder ce n'est pas (forcément) réutiliser !

## - Licences Creative Commons

## - Licence ouverte Etalab

Publication des données publiques



ATTRIBUTION	
ATTRIBUTION / PAS DE MODIFICATION	+
ATTRIBUTION / PAS D'UTILISATION COMMERCIALE / PAS DE MODIFICATION	+  +
ATTRIBUTION / PAS D'UTILISATION COMMERCIALE	+
ATTRIBUTION / PAS D'UTILISATION COMMERCIALE / PARTAGE DANS LES MÊMES CONDITIONS	+  +
ATTRIBUTION / PARTAGE DANS LES MÊMES CONDITIONS	+



# Diffuser ses données

*Les questions à se poser pour alimenter le PGD*

- **Quelles données conserver et diffuser parmi toutes celles produites ou collectées et pourquoi ?**
  - Données brutes ? Données structurées ? Et pourquoi ?
  - Intérêt scientifique, économique, démocratique, environnemental, social ?
  - Dans quels délais ? Embargo ?
  - Quelles données ne le seront pas et pourquoi ?
- **Quel est le potentiel de réutilisation des données ?**
  - Quelles seraient les réutilisations possibles ?
- **Avez-vous des obligations ?**
  - Suppression ou conservation ? (obligation de suppression des données personnelles après la recherche, base de données...)
- **À qui appartiennent les données ?**
  - Est-ce qu'une propriété privée s'exerce sur les données ?
  - Obtenir toutes les autorisations pour diffuser



# Diffuser ses données

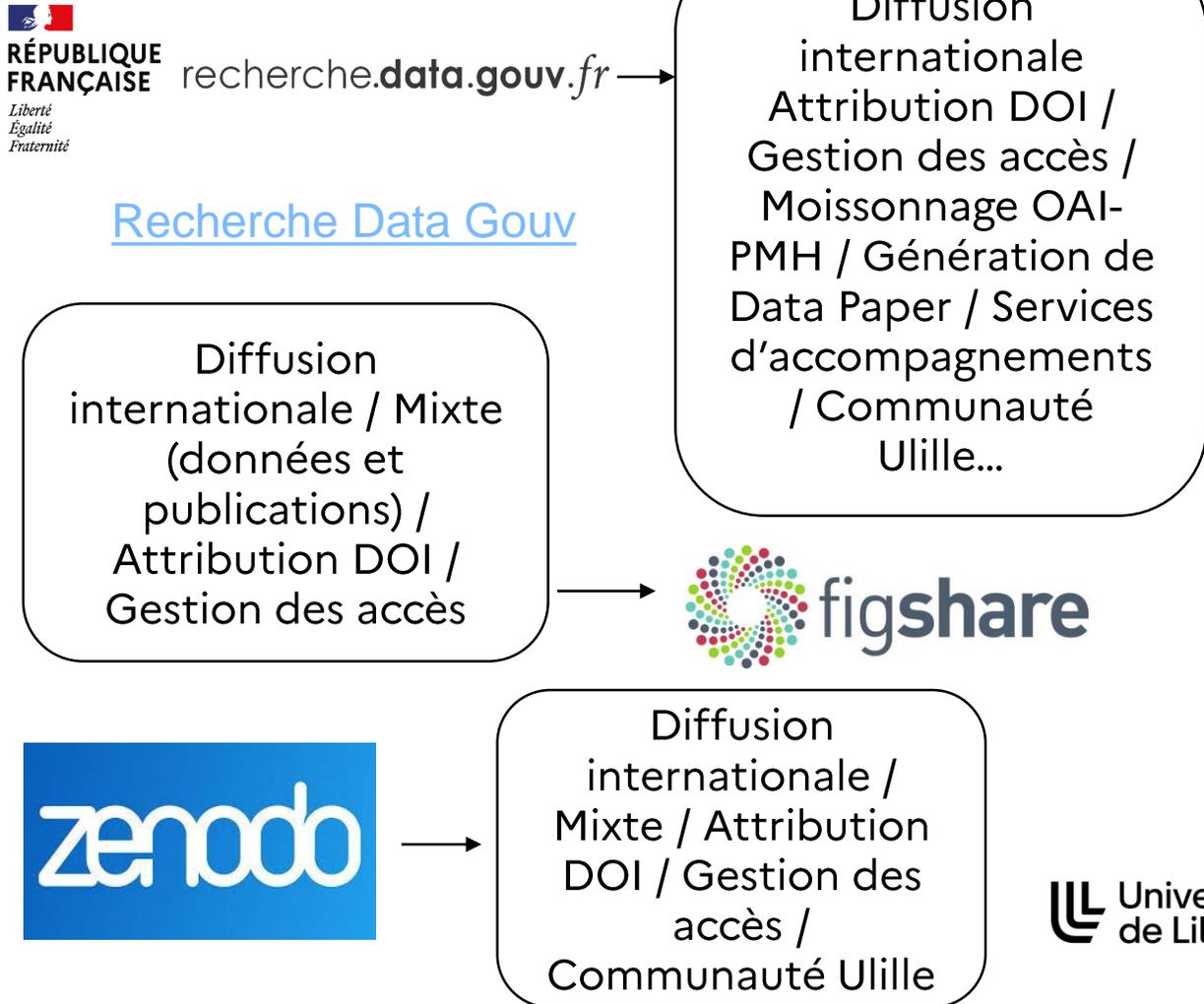
## *Les questions à se poser pour alimenter le PGD*

- **Est-ce qu'il y a des données sensibles ?**
  - Ne pas diffuser de données à caractère personnel, secret-défense, secret industriel, etc.
  - Diffuser des données en accès restreint quand nécessaire (restreint à certains publics, disponibles sur demande justifiée...)
- **Quel format de données ?**
  - Privilégier les formats ouverts non propriétaires
- **Quelle méthode sera utilisée pour garantir le respect des principes FAIR ?**
  - Un identifiant pérenne sera-t-il attribué aux données ? Lequel ?
  - Quel sera l'entrepôt de confiance utilisé pour la diffusion des données ?
- Faire des choix participe à la sobriété numérique car cela permet d'éviter de conserver et de partager des données inutiles ayant un impact environnemental conséquent



# Comment diffuser vos données ?

## Entrepôts généralistes



## Entrepôts disciplinaires



# Comment diffuser vos données ?

## Un entrepôt est-il préconisé par :

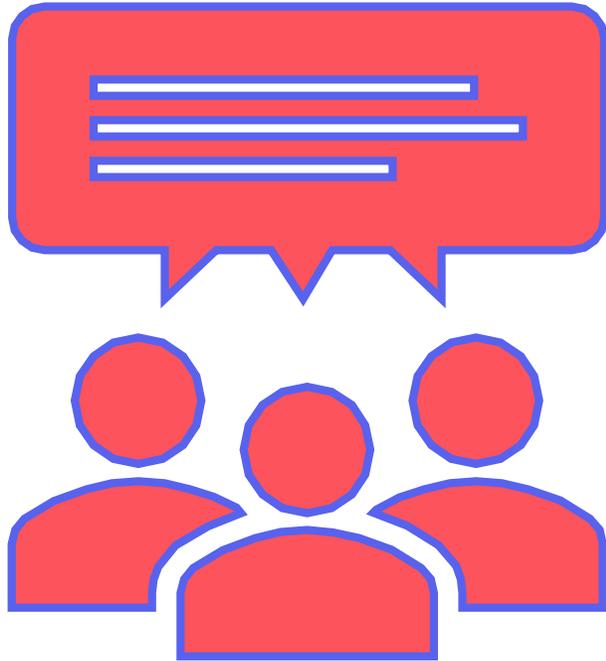
- La discipline ?
- L'institution de recherche ou partenaires ?

## Trouver un entrepôt de données grâce à des annuaires spécialisés :

- Re3data : <https://www.re3data.org/> (moteur de recherche des entrepôts de données qui affine les résultats en fonction des besoins).
- CatOpidor : <https://cat.opidor.fr/index.php/Conservation> (permet d'identifier les services nationaux dédiés aux données de recherche)

## Critères pour choisir un entrepôt :

Discipline	Droits d'accès	Identifiant pérenne	Versioning	Certification (Core Trust Seal)
Métadonnées adaptées	Format	Propriété des données	Volumétrie	Liaison entre données et publications
Lieu physique de stockage (politique de l'entrepôt et du pays en vigueur)	Durée de conservation	Coûts	Statistiques d'utilisations	



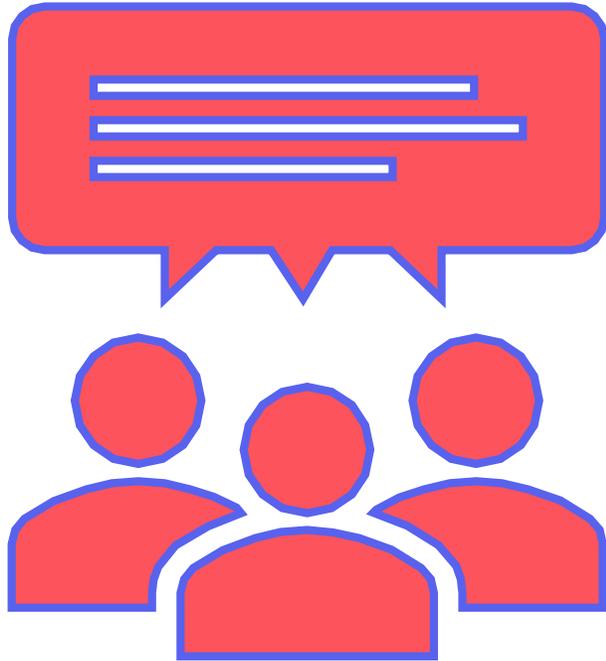
## 6 Responsabilités et ressources

# Responsabilités et ressources

*Les questions à se poser  
pour alimenter le PGD*

- **Qui est responsable de la gestion des données ?**
  - Décrire les rôles et responsabilités
  - Partenaires ?
  - Nom des personnes selon les étapes du cycle de vie des données
  - Personne responsable du plan de gestion de données
  - Personne en charge de la coordination de la rédaction, mise à jour, mise en œuvre...
- **Quelles ressources seront dédiées à la gestion des données et permettront de s'assurer de les rendre « FAIR » ?**
  - Temps de travail: Qui ? Pour quelles tâches ?
  - Frais de stockage: coût, répartition entre partenaires
  - Coûts de mise à disposition des données: entrepôts, anonymisation, préparation des données





## Aide à la rédaction

# Le Service commun de documentation accompagne le montage et le suivi de vos projets ANR et européens

## Coordinateurs de projets ANR et européens financés :

- ❖ Ateliers sur les plans de gestion de données
- ❖ Relecture commentée des plans de gestion de données
- ❖ Administration de Recherche Data Gouv
  - ❖ Aide au dépôt des données
  - ❖ Aide à la création de collections pour le projet
  - ❖ Formation sur l'utilisation de l'entrepôt
- ❖ Ateliers « respect des obligations de diffusions des publications en open access »

## Intégrer la science ouverte dans le montage des projets ANR et Horizon Europe

- ❖ Accompagnement au montage pour les projets européens et ANR
- ❖ Ateliers collectifs AAPG ANR

Marie Cros  
Chargée de projet Données

Mickael Malandran  
Coordinateur Archives Ouvertes et chargé de projet Données

**Pour en savoir plus :** <https://bu.univ-lille.fr/chercheurs-doctorants/science-ouverte/donnees-de-recherche>



**Contact :** [scd-aap@univ-lille.fr](mailto:scd-aap@univ-lille.fr)



# la Fabrique de la science ouverte

23 mai - 21 juin 2023



Le Printemps  
de la Donnée

## Programme

*Les bibliothèques universitaires et Learning center de l'Université de Lille vous proposent, dans le cadre du printemps de la donnée, du 23 mai au 21 juin 2023 :*

en distanciel

23 mai | Les principes F.A.I.R., pourquoi et comment faire ?

26 mai | Atelier Plan de Gestion des Données express

30 mai | Découverte de Recherche Data Gouv

en présentiel à LILLIAD Learning center Innovation (salle Y)

+ en distanciel

06 juin | Les forges universitaires au coeur de l'ouverture des logiciels de la recherche

21 juin | LORD : Les acteurs de la donnée ULille

Informations et inscriptions : [fabso.univ-lille.fr](https://fabso.univ-lille.fr)

À noter dans son agenda  
en automne : l'Open Access Week  
(plus d'infos à venir).