

# De la gestion de ses données au dépôt de la thèse

Journée de rentrée ED CESCO – 20/10/2025

Séverine Ouane (Service Appui à la Science Ouverte / SCD)



## Gestion des données de la recherche : de quoi parle-t-on ?

---

*Imaginons une équipe de recherche qui s'intéresse au développement du bilinguisme chez les enfants issus de familles franco-allemandes ;*

*elle travaille à partir de sources variées : questionnaires remplis par les parents, enregistrements audio des enfants racontant une histoire dans chaque langue, observations du quotidien linguistique, informations sur les activités en français et en allemand, tests de comportement social et affectif...*

*Pour chaque enfant étudié, les chercheurs notent l'âge, la quantité d'exposition à chaque langue, les personnes avec qui il parle français ou allemand, les activités réalisées dans chaque langue, ses attitudes envers ses langues, son aisance à s'exprimer, ainsi que des données sur son bien-être émotionnel et social.*

*Petit à petit, cette collecte devient une base de données structurée, permettant d'explorer comment le bilinguisme se développe selon les profils familiaux, la qualité de l'input linguistique, les attitudes des enfants ou encore leur équilibre socio-affectif.*

## Une donnée de la recherche : c'est quoi ?

---

*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 des résultats de recherche.*

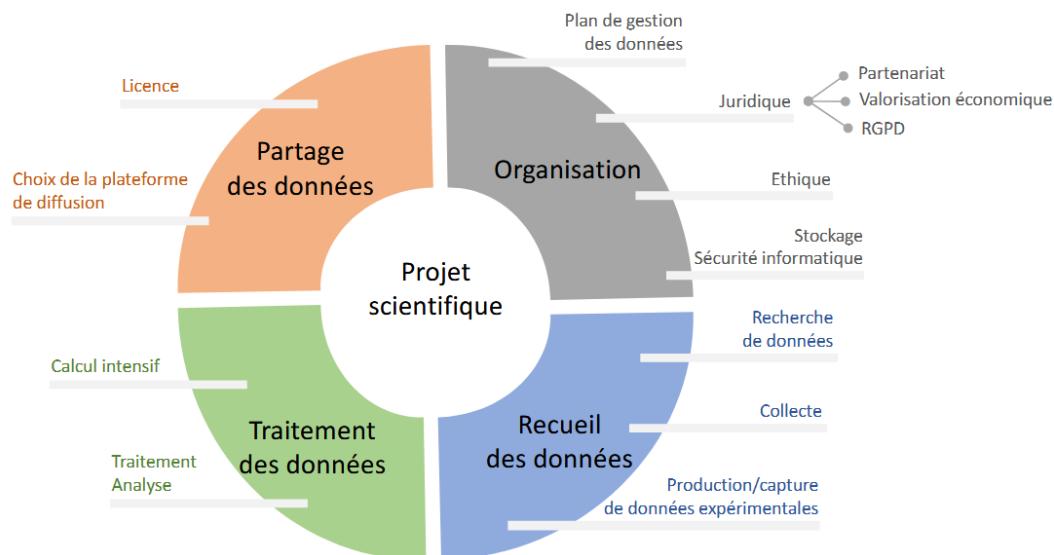
Définition proposée par l'OCDE dans son rapport « [Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics de 2007](#) ».

*Ici :*

- *un enregistrement audio*
  - *une transcription des entretiens*
  - *Des résultats aux questionnaires*
  - *une mesure de la durée moyenne des syllabes*
  - *une longueur moyenne des séquences*
  - *un emplacement des pauses*
  - *un nombre de disfluences par minute*
- = des données issues de sources hétérogènes, combinées et structurées ensemble pour être analysées.*

# Gestion des données de la recherche : de quoi parle-t-on ?

---



*Titularité des données  
Droit de (ré)utilisation  
Données personnelles  
ou sensibles*

### À qui appartiennent les données ?

Une doctorante passe plusieurs mois à renseigner une base de données. Un ingénieur a développé un outil de saisie. Le projet est cofinancé par un laboratoire et une clinique privée.

- Est-ce que la doctorante peut réutiliser les données pour un autre projet après sa thèse ?
- Est-ce que la clinique peut demander à publier une synthèse sans l'équipe ?

Cette question soulève la nécessité d'un accord clair dès le départ sur la titularité des données produites collectivement.

## Gestion des données de la recherche : de quoi parle-t-on ?

---

### **A-t-on le droit de réutiliser ce qu'on collecte ?**

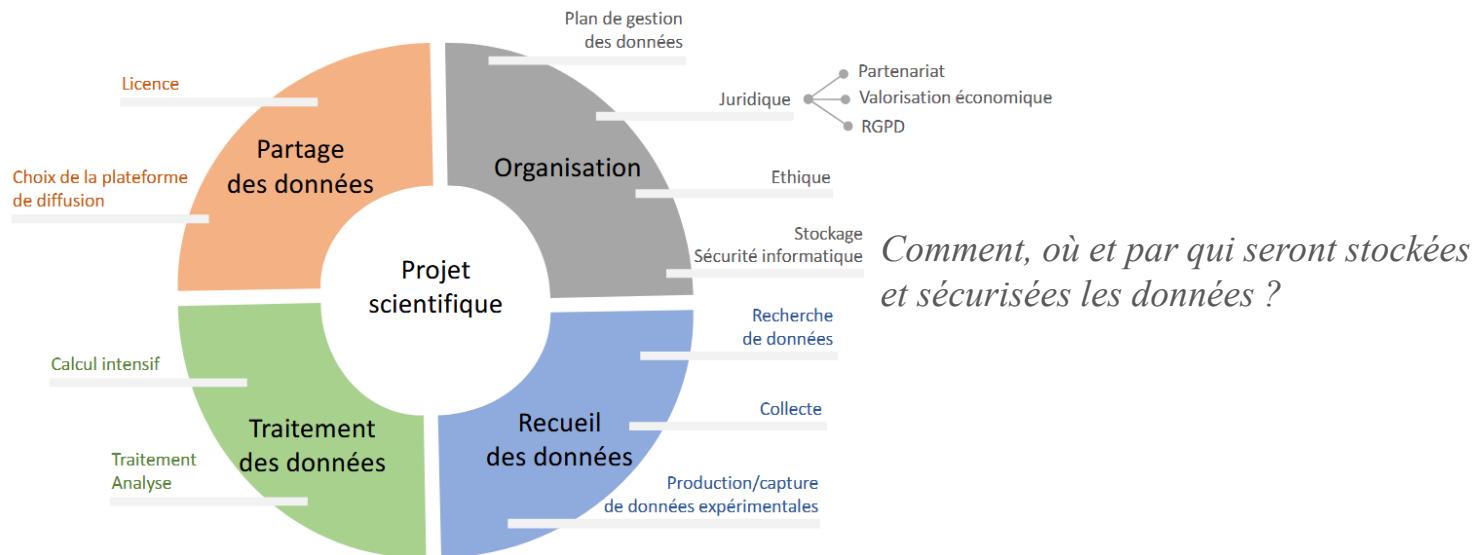
L'équipe souhaite utiliser des données générées dans des projets antérieurs (par exemple des enregistrements d'enfants ou des réponses à des tests psychologiques).

- Quels sont les droits rattachés à chacune des données ?
- Si il y a une possibilité de consultation, existe-t-il la possibilité de les réutiliser dans une nouvelle étude?

Il faut vérifier systématiquement les conditions de réutilisation pour éviter les conflits juridiques.

# Gestion des données de la recherche : de quoi parle-t-on ?

---



## Gestion des données de la recherche : de quoi parle-t-on ?

---

### Comment et où les données seront-elles stockées ?

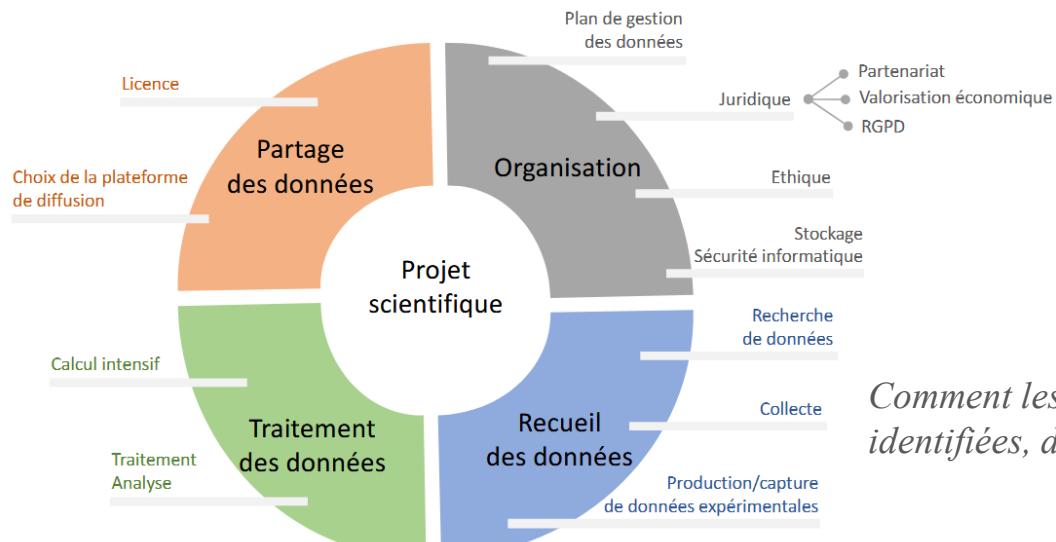
Un fichier Excel est sur l'ordinateur d'un des chercheurs. Une autre version circule par mail. Un doctorant travaille sur une copie, avec des modifications.

- Qui a la bonne version ?
- Que se passe-t-il si l'ordinateur plante ?

Il faut une stratégie de stockage partagé, sécurisé, sauvegardé. Cela peut passer par une plateforme institutionnelle, un entrepôt sécurisé ou un outil collaboratif bien géré.

# Gestion des données de la recherche : de quoi parle-t-on ?

---



*Comment les données seront identifiées, décrites ?*

## Gestion des données de la recherche : de quoi parle-t-on ?

---

### **Comment les données seront-elles décrites, identifiées ?**

La variable « durée moyenne des syllabes » est parfois notée en secondes, parfois en millisecondes,...

Difficile de faire des analyses comparées si les unités de mesure varient.

Il faut penser standardisation, métadonnées, vocabulaire contrôlé (même artisanal au début), pour que les données soient lisibles, exploitables et comparables dans le temps.

# Gestion des données de la recherche : de quoi parle-t-on ?

---

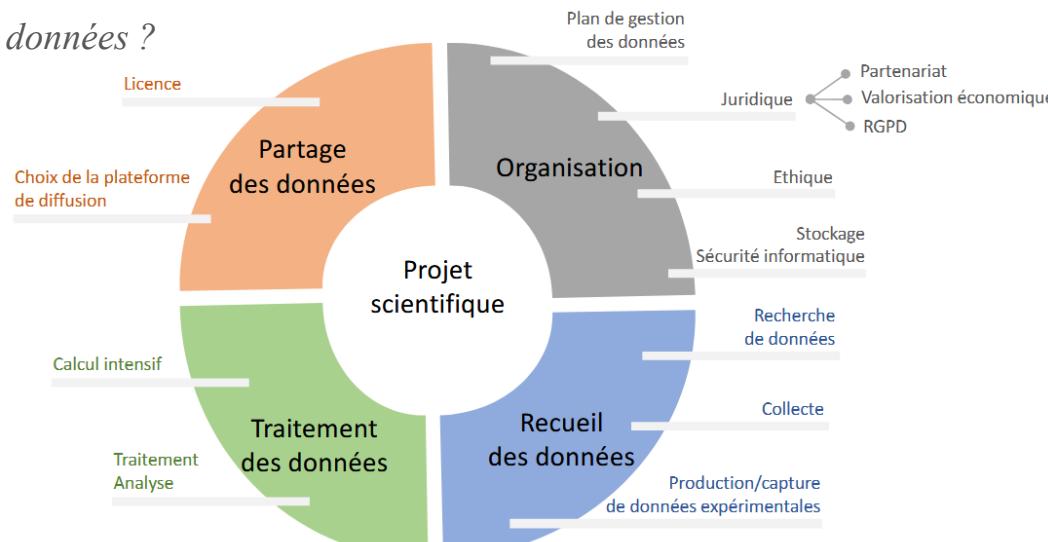
*Qui pourra accéder aux données ?*

*Seront-elles diffusées ?*

*Comment ?*

*Quand ?*

*Sous quelle licence ?*



## Gestion des données de la recherche : de quoi parle-t-on ?

---

### **Qui a accès aux données, et à quel moment ?**

Une stagiaire en master voudrait exploiter une partie de la base.  
Un partenaire du projet à l'étranger demande à y accéder.

- Tout ouvrir d'un coup peut poser des problèmes.
- Refuser l'accès peut bloquer la dynamique de projet.

Il faut définir des droits d'accès différenciés : en interne, en externe, temporairement ou durablement, selon les rôles.

## Gestion des données de la recherche : de quoi parle-t-on ?

---

### **Les données seront-elles diffusées ? Quand ? Comment ?**

L'équipe souhaite créer une plateforme interactive pour visualiser les profils bilingues des enfants. Mais certains éléments sont encore en cours de vérification.

- Doit-on attendre la fin du projet ?
- Peut-on publier les données validées progressivement ?

Il faut planifier la stratégie de diffusion : en une ou plusieurs étapes, via un dépôt (type Nakala).

## Gestion des données de la recherche : de quoi parle-t-on ?

---

### **Sous quelle licence les données seront-elles mises à disposition ?**

Un fichier de synthèse pourrait être utile à d'autres chercheurs, mais l'équipe veut être correctement citée.

- Licence ouverte avec citation obligatoire (CC BY) ?
- Autorisation de toute réutilisation (CC0) ?
- Interdiction d'usage commercial ?

Choisir une licence claire, c'est éviter les malentendus, et encourager une réutilisation respectueuse du travail réalisé.

## Gestion des données de la recherche : pourquoi s'en préoccuper?

---

### 1. Pour soi-même

- Retrouver ce qu'on a fait deux ans plus tôt
- Structurer son travail pour mieux l'exploiter
- Reprendre une base, une analyse, une source

### 2. Pour les autres

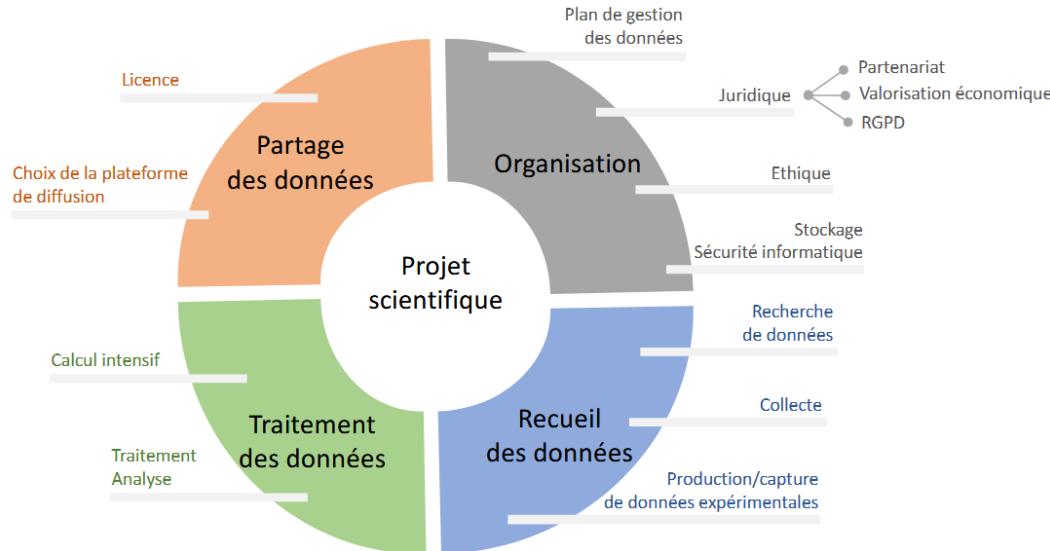
- Travailler en équipe, transmettre à un doctorant
- Permettre la réutilisation scientifique (et les citations)
- Éviter que tout repose sur une personne ou un PC

### 3. Pour répondre aux exigences

- Certains financeurs (ANR, Europe) demandent un plan de gestion des données (PGD) et l'ouverture des données
- Les revues ou éditeurs demandent de plus en plus le dépôts des données associées aux publications
- On parle de plus en plus de science ouverte : les données peuvent être des résultats en soi.

# Gestion des données de la recherche : le Plan de Gestion de Données

*Check-list de points d'attention > feuille de route tout au long du projet*



## Plan de Gestion de Données : Pourquoi le faire dès le début du projet ?

### Clarifier dès le départ ce qu'on va produire comme données

→ Évite les surprises, permet de dimensionner les moyens nécessaires

### Identifier les besoins techniques, humains et financiers

→ Stockage, outils, compétences, accompagnement, budget

### Prévoir les contraintes avant qu'elles ne bloquent

→ Propriété intellectuelle, éthique, RGPD, obligations des financeurs

### Attribuer des rôles et responsabilités

→ Qui fait quoi ? Qui documente ? Qui décide du partage ou de l'archivage ?

### Choisir des formats et des standards adaptés dès le départ

→ Garantit la qualité et la réutilisabilité des données

### Éviter les pertes de temps et les bricolages en cours de route

→ On construit une base solide sur laquelle le projet peut avancer sereinement

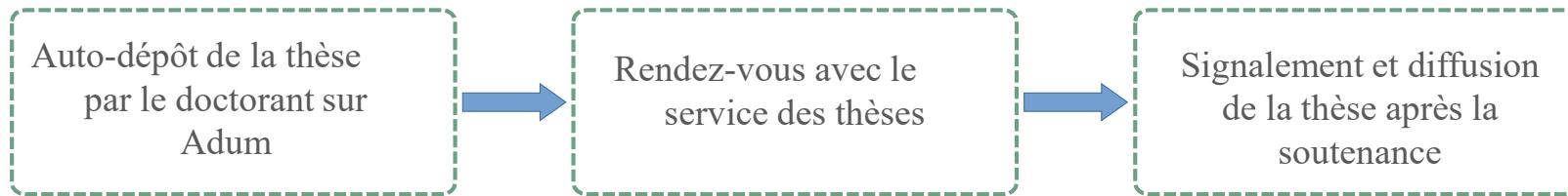
## Gestion des données de la recherche : ce qu'on peut vous apporter

---

- Aide lors de la rédaction des réponses aux appels à projet (frais liés à la gestion des données sont éligibles à remboursement)
- Appui à la l'élaboration du Plan de Gestion des Données (PGD)
- Appui à la mise en place de bonnes pratiques en lien avec les différents acteurs de l'établissement
- Appui à la description des données
- Accompagnement au dépôt des données

Contact : [dmp@univ-tlse2.fr](mailto:dmp@univ-tlse2.fr)

## Gestion des données de la recherche : ce qu'on peut vous apporter



- Liens utiles
- [Page des bibliothèques consacrée aux thèses](#)
- [Formations aux doctorants](#)
- [Plan de gestion des données](#)